NEW METHOD FOR MATHEMATICAL MODELLING OF HUMAN VISUAL SPEECH

MOHAMMAD HOSSEIN SADAGHIANI, MSc.

Thesis submitted to the University of Nottingham

for the degree of Doctor of Philosophy

July 2015

ACKNOWLEDGMENT

In memory of my father,

To my mother,

To my Uncle Majid who encouraged me during my study.

The author would like to appreciate the help of Professor Ian Harrison the Dean of the engineering faculty at the University of Nottingham Malaysia campus in proof reading the contents of this thesis.

Contents

FIGURES

TABLES

ABBREVIATIONS

BLI	Barycentric Lagrange Interpolation
BLP	Barycentric Lagrange Polynomial
FP	Feature Points
BDC	Barcode Duty Cycle
$P(.)$	Polynomial
p	Point
C	Consonant
V	Vowel
c_v	Constant Value
$a, b, c, \dots$	Scalars, constant coefficients: lowercase letters
$\|a\|$	Length of a, absolute value of a
x, y, z	Cartesian coordinates
$a \in b$	a was an element of b
$a \approx b$	a was almost equal to b
$f(.)$	Function
$[a, b]$	Closed interval, continuous values
$[a: b]$	Closed interval, discrete values
$\mathbb{N}$	Natural numbers, Integers: boldface, uppercase
u, t, i, j	Parametric variables
f_i	Frame sequence (discrete-time domain)
f_x	Frame sequence (continues-time domain)
F	Number of frames
T	Sampling interval
$p(x_{x^i}, y_{x^i})$	Pixel coordinate
t^f	Frame position in time domain
M, K	The ROI's length and width
W	Word
F_W	Word's frame length
S_W	Word's time duration
u	Upper
l	Lower
c	Corner
u^W	Difference vector $(y_{t^f}{}^c - y_{t^f}{}^u)$
l^W	Difference vector $(y_{t^f}{}^l - y_{t^f}{}^c)$
c^W	Difference vector $(x_{t^f}{}^c - x_{t^f}{}^u)$
$F_{W_m}^u$	The upper feature point
$F_{W_m}^l$	The lower feature point
$F_{W_m}^c$	The corner feature point
$FP_{V_m}^{u'}$	The extracted upper feature point
$FP_{V_m}^{l'}$	The extracted lower feature point
$FP_{V_m}^{c'}$	The extracted corner feature point
$VS_{BLI}^u{}^{W_m}(f_x)$	Upper visual speech signal
$VS_{BLI}^l{}^{W_m}(f_x)$	Lower visual speech signal
$VS_{BLI}^c{}^{W_m}(f_x)$	Corner visual speech signal
β	Sampling factor
n	Integers

$I_s(.)$	Interpolated signal
x_d	Decimated signal
x_u	Expanded signal
φ	Basic functions
Φ	Schoenberg's Basic function
$l(.)$	Lagrange basic functions
$w(.)$	Barycentric weighting function
e	Error
$tri(.)$	Triangle pulse train
τ	Bar's width
$F_{2D}(.)$	Two dimensional scatter function
L_q	Quantization level
s_q	Symbol
$I(.)$	Information
H	Entropy
ω_0	Angular frequency
f_0	Fundamental frequency

ABSTRACT

Audio-visual speech recognition and visual speech synthesisers are used as interfaces between humans and machines. Such interactions specifically rely on the analysis and synthesis of both audio and visual information, which humans use for face-to-face communication. Currently, there is no global standard to describe these interactions nor is there a standard mathematical tool to describe lip movements. Furthermore, the visual lip movement for each phoneme is considered in isolation rather than a continuation from one to another. Consequently, there is no globally accepted standard method for representing lip movement during articulation. This thesis addresses these issues by designing a transcribed group of words, by mathematical formulas, and so introducing the concept of a visual word, allocating signatures to visual words and finally building a visual speech vocabulary database. In addition, visual speech information has been analysed in a novel way by considering both lip movements and phonemic structure of the English language. In order to extract the visual data, three visual features on the lip have been chosen; these are on the outer upper, lower and corner of the lip. The extracted visual data during articulation is called the visual speech sample set. The final visual data is obtained after processing the visual speech sample sets to correct experimented artefacts such as head tilting, which happened during articulation and visual data extraction. The 'Barycentric Lagrange Interpolation' (BLI) formulates the visual speech sample sets into visual speech signals. The visual word is defined in this work and consists of the variation of three visual features. Further processing on relating the visual speech signals to the uttered word leads to the allocation of signatures that represent the visual word. This work suggests the visual word signature can be used either as a 'visual word barcode', a 'digital visual word' or a '2D/3D representations'. The 2D version of the visual word provides a unique signature that allows the identification of the words being uttered. In addition, identification of visual words has also been performed using a technique called 'volumetric representations of the visual words'. Furthermore, the effect of altering the amplitudes and sampling rate for BLI has been evaluated. In addition, the performance of BLI in reconstructing the visual speech sample sets has been considered. Finally, BLI has been compared to signal reconstruction approach by RMSE and correlation coefficients. The results show that the BLI is the more reliable method for the purpose of this work according to Section 7.7.

Chapter 1

INTRODUCTION

In this thesis, modelling the visual speech information by explicit mathematical expressions coupled with words' phonemic structure analysis has been represented. These expressions have been derived based on manual extraction of visual speech data. This is due to the fact of inaccurate automatic visual data extraction in comparison with manual visual data extraction (see Section 4.5). The visual information has been obtained from deformation of lip's dimensions during articulation of a set of words.

1.1. VISUAL SPEECH INFORMATION

The invention of telephone had upgraded the level of human communication in the end of 19th century. After almost 4 decades, the scientific analysis and synthesis of the audio signal started in 1930 for speech processing (Dudley, et al., 1939). In the mid-20th century, the demands for more sophisticated interfaces caused the usage of the other perspective in communication. Gradually, the need for interacting with machine also appeared in human society. Inspired by human communication via speech and facial clues, the computers can be trained to use same abilities. The usage of the speech information in form of lip movements started around two decades earlier than the visual information. One of the main reasons for such delay between audio signal processing and visual speech signal processing was the computational cost of visual speech due to its dimensions. In other words, the speech signal has one dimension while its visual data could have two or three dimensions adding higher analysis and synthesis complexities.

In early 1970, computer data processing speed as well as its memory reached to a level suitable for visual data interpretations and simulations. The first animated talking head introduced by a novel work of Park (Park, 1972). Since then a great effort for employing such ability in technology was begun. It has been shown that the visual cues of speech also enhance the transparency of speech when it was degraded by noise (Sumby & Pollack, 1954) since they are more robust to environment noise.

Studying the visual information is started by relating the smallest speech component (phoneme) to the dynamism of lip during articulation. Therefore, in the visual domain, a basic component related to phoneme defined (Fisher, 1968) and called visemes (visual-phonemes). The observed visual phonemes are categorized using confusion matrix (Williams, et al., 1997) where it is the most accurate phoneme form a phoneme-viseme table.

The visual speech information is a bimodal component of audio speech signal (Fisher, 1968). These two modalities are strongly correlated. The perception of lip movement can be affected in presence of different audio speech signal (McGurk & MacDonald, 1976).

The applications of human visual information are considered as improvement for degraded audio speech, lip-reading systems (Petajan, 1984), and animation (Parke, 1982). Besides, the ability of separating different audio speech sources using visual speech information was demonstrated by (Girin, et al., 2001).

Basically, there are two varieties of Human-machine interactions. In the first category, the machine plays a role as interface in human-to-human interaction (audio-visual speech recognition) or animation (visual speech synthesis). The other possibility is machine-to-machine interaction with human standards of communication (artificial intelligence). In all cases, the necessity of accurate perception and generation of human communicating abilities as well as standard interpretation of them are desirable.

In the audio-visual automatic speech recognition (AVASR) systems, the visual information is the input of identification for enhancing the perception of speech signal modality. This visual recognition module is interchangeably called lip-reading, visual

speech recognition (Petajan, 1984), speech reading, or visual-only automatic speech recognition (Potamianos, et al., 2004).

In the lip-reading systems, by focusing on the appearance of the lip's geometry or pixels colour in image sequences, the dynamic of articulating lip are extracted. This information will be processed in order to be compatible for fusing to the audio-only ASR modality. The usage of viseme appears in this stage. On the other hand, in animation or visual speech synthesis the speech signal (Massaro, et al., 1999) or transcribed information of words (Ezzat & Poggio, 1998) are used for lip's movement identification. In both cases, the lip movement is interpreted from phoneme-viseme table.

Furthermore, in AVS processing, the visual expressions can be used as audio speech anticipators. The anticipation effect was represented in (Kim & Davis, 2003) where in (Conrey & Pisoni, 2003) the asynchrony of visual speech and audio speech information perception was examined. It has been shown that the tolerance of perceivers to visual speech cues preceding the audio speech cues was higher.

The main issue in viseme concept is lack of globally accepted definition for relation between phoneme and its visual appearance. The next drawback of the viseme concept is the lack of accuracy in perception of visual phonemes due to dependency on human perception of visual phonemes. In other words, depending to the observers whether they are hearing impaired or not, the results are changed.

The next important issue is lack of explicit definition of the lip's deformation to make a standard model of articulating lip. The concept of visual phoneme does not suggest an explicit definition of lip's structure during phoneme utterance. In addition, the main factor of human perception in evaluating and selecting viseme could vary from one observer to another. Therefore, there is no global agreement on standardized phoneme-viseme table reference.

Furthermore, the sequential relation of visemes totally changes the shape of isolated visemes. In other words, combination of visemes does not satisfy the realism of visual lip movement. This fact in visual domain is tied to the Co-articulation effect that considers the effect of phonemes on each other in a sequence. Although, the Co-

articulation model in visual domain tackles the issue in the field of visual speech synthesis (Cohen & Massaro, 1993), but the countless combinations of phonemes remains problematic. In this work, the proposed approach has solution for the doubts about the viseme concept and Co-articulation effect in definition of the visual phonemes.

The suggested solution begins with analysis of linguistic rules in phonemic level, which then used for mathematical formulation of them in visual domain. Furthermore, the mathematical expressions of articulating lip are emerged from phonemic rules. In the first stage phonemic structure is added to articulated words. Then, the mathematical expressions unify the lip's movements. It is obvious that all of phoneme combinations do not necessarily reveal meaningful words. It is also clear that even all phoneme combinations are provided, finding the meaningful ones are very tedious and time consuming. Even if these combinations are found, the Co-articulation effect still would be needed to apply to phonemes. On the other hand, the proposed model of analysing the phonemic structures related to the words phonemic structure preserves Co-articulation. Furthermore, this approach creates a systematic relation between words. The relation emerges from the similarity of phonemes in the beginning of words. For example, the word 'circumstance' and 'circle' share a similar audible data in the beginning. Such similarity also appears in the visual domain during uttering these two words. These words have the same root but as they progress, different phonemes appear. This concept could be applied to a population of words such as dictionary by finding how phonemes progress and succeed by others. After finding these relations as rigid representations of words, their visual modalities, the visual speech concept obtain a new dimension different from what is suggested in viseme concept.

The next challenging issue could be raised about the accuracy of generated speech and realism of the lip movement. A reference of visually perceived language, which is ideally accepted, must be defined. In the English language, there is no commonly accepted reference for articulating the words. For instance, the variation of the manner of articulation is different in Australia to the New Zealand. Therefore, one of the future challenges is defining a reference, which is accepted globally. For other language such as French or Spanish, it is possible to find standards that could be

accepted in their regions by linguists. The English language is chosen to be applied to the proposed method of visual speech formulation in this work.

One of the important applications of proposed method is the English language education. The other is for formulating visual speech information in the field of lip-reading. Instead of using the probabilistic analysis of speaker's lip movement, the lip-reading systems would be equipped with mathematical expressions of visual phonemes in their database. Therefore, more efficient decision on detecting the words could be made. Such efficiency emerges from the characteristic of defined words as they share the same roots. By tracking the beginning of words, in specific times, the number of possible words would be reduced resulting in higher efficiency and avoiding unnecessary calculations.

In the field of visual speech animation, the formulated visual expressions could be a reliable successor of viseme mapping scheme. The expressions could be chosen by animator and bind to the controlling parameters of lip model (template). The advantages of visual speech expressions make them a reliable replace in language education, lip-reading and visual speech animation systems.

The focus of this thesis is on modelling of the visual speech information via a mathematical tool. The model constructs the dynamic of lips from its geometrical structure and allocates the concept of visual speech signal to the lip's movement. The relation of visual speech signal and its speech signal are built based on a novel suggestion to organize words based on phonemic structures. When the words are categorized by a particular phonemic rule, the unity between visual speech signal, speech signal and the transcription of words is guaranteed. Preserving the Co-articulation effect is another advantage of such modelling by eliminating the usage of any function that defines a transition pattern between the successive visual phonemes. The suggested method is the first step toward convergence of all visual speech analysis and synthesis methods used in AVASR and visual speech synthesisers to a globally standardized approach. The constructed visual speech signal is compared with another signal construction method.

1.2. Problem statement

As mentioned, there are three main issues in the fields of visual speech recognition and animation. These issues are related to interpreting the lip movements during articulation of isolated phonemes. First, instead of representing the visual phoneme sequence (a word) explicitly, it has been explained by non-standard group of lip frames or the statistical characteristics of lip's geometry or pixel colour. Second, preserving the continuity of visual phoneme sequences (Co-articulation effect) has been suggested by complementary methods. In addition, there is no reported standard method which has been globally accepted for representing the lip movement.

1.3. Aim

This thesis concentrates on the visual speech components and introduces a novel method of modelling the meaningful combinations of phonemes using linguistic rules in a limited numbers of transcribed words. The foundation of this model consists of relating the mathematical models, visual speech information, and transcribed word dictionary to gather.

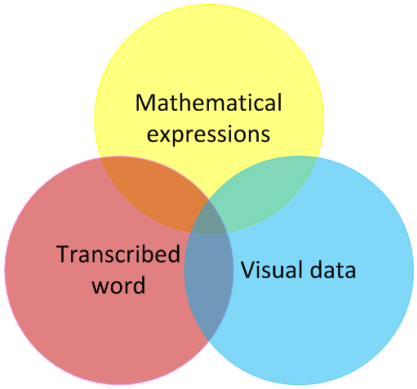


Figure 1-1: The core of the thesis

In Figure 1-1, the union of the three concepts used depict the foundation of this thesis. The outcomes of such relations are:

- Analysing a transcribed word corpus, systematically
- Deriving mathematical expressions for corresponding lip movement based on the transcribed word corpus
- Standardising the visual speech information

The mathematical expressions are derived for the lip movements during articulation of a systematically designed transcribed word corpus. In order to standardized the mathematical expressions, further identification properties are dedicated to the visual speech information.

1.4. OBJECTIVE

More specifically, in order to justify the lip movement can be represented with mathematical expressions; the objectives of this work are:

1. Designing the corpus which has been chosen from an automatic speech recognition dictionary. The corpus has been analysed by finding a linguistic rule in the English language via hierarchical analysis. The result of analysis is a group of transcribed words that has the same three phonemic sequences in their beginnings.
2. Deriving the mathematical lip movement expression by BLI from extracting three visual speech feature point on the upper, lower and corner of speakers lips.
3. Defining the visual words by combining the mathematical expressions of lip. More specifically each visual word has:
    - Mathematical expressions
    - Visual representation
    - Barcode representation
    - Digital and encoded representations
    - Two/three dimensional representations
    - Volumetric representation (volumetric visual word)
4. Defining the signatures for the visual words by allocating mathematical expressions using BLI, barcode representations using PWM, digital representations (quantized and encoded visual words), 2D/3D visual words and volumetric modelling of the visual words.
5. Constructing a visual vocabulary database.

The practical phase of this work is shown in Figure 1-2, which will be described with more details.

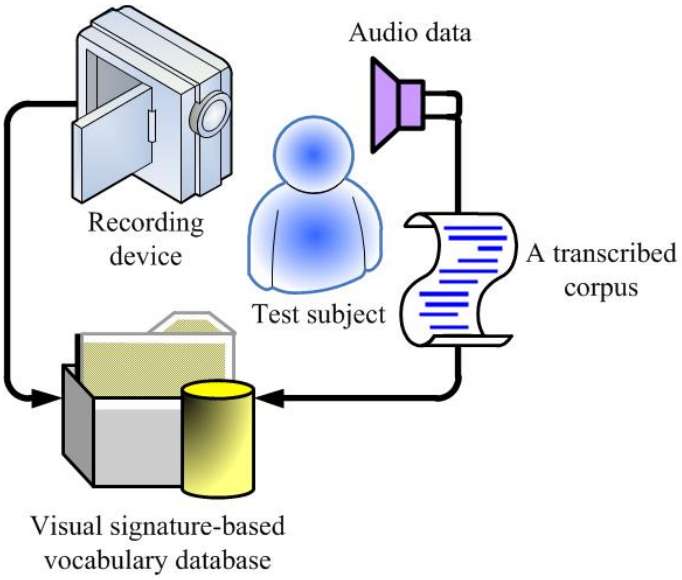


Figure 1-2: The practical phase of the work

The visual speech signals are processed for signature allocation to words. The words are categorized based on the hierarchical relations in phoneme sequences. A word is represented in visual domain as visual word with specific mathematical expressions and signatures. The mathematical signature of a visual word consists of three visual speech signals. These visual speech signals will be interpolated from three static visual features located on the upper, lower and corner of outer lip contour. The mathematical expressions are derived by the Barycentric Lagrange Interpolation method. The number of visual speech signals was set to three since the three visual features include the necessary information of lip dimension during articulation. The reason of using BLI is the ability of formulating the visual speech signals which involve the sample sets extracted from the visual feature points without Runge's effect. The mathematical signatures themselves have other subgroups. These signatures are called as visual word barcode, digital visual word, and visual word 2D/3D signatures. These three types of signatures are configured as signature database. The first type concludes the signature signals (polynomials) by barcode allocation. The second type uses a coding scheme to form coded version of signature signals. By relating the signature signals to each articulating lip in two or three dimensions, the third subgroup of signature signals is obtained. Finally, by allocating a lip template to the visual speech signals for each word in designed corpus, volumetric representation of lip's movement are also suggested.

In addition to the mathematical signatures, a lip template has been designed for animating the visual words. In order to compare the results, the expressions are also generated using signal reconstruction method by the sinc function. The results of comparison are represented in form of Root Mean Squared Error (RMSE) and correlation coefficients. An overview of the main stages in modelling the human visual speech is shown in Figure 1-3.

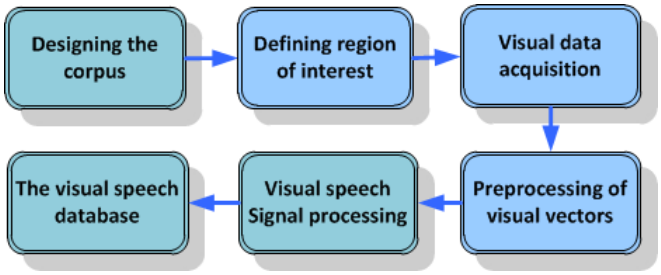


Figure 1-3: An overview of modelling the human visual speech

After designing a corpus, the subjects are asked to articulate the words. Their lips' movements are captured by a camera. The lips geometry is parameterized on the region of interest by three static feature points located on upper, lower and corner outer contour of lip (see Section 4.1). Therefore, each word has three sets of samples in visual domain. The movement of these feature points can describe the lips' movement. The visual data is extracted automatically and manually (see Section 4.4). The acquired data needs to be processed in the words to reduce the errors that have been introduced by the act of articulating as well as feature points extraction. Using the Barycentric Lagrange Interpolation (BLI) leads to represent the lip's movement trajectory by three visual speech signals. These polynomials are called visual speech signals.

An evaluation of the interpolation's performance on three different types of data has been conducted (Chapter 6). The interpolation process provides a polynomial including the data samples. Such polynomial can also be considered as signals. The polynomials further process leads to three subgroups of signatures.

1.5. NOVELTIES OF THE WORK

In this thesis, the suggested novel approaches for tackling the issues in the visual speech domain are:

- A new approach for designing transcription of words (corpus) based on phonemic rules.
- Expressing the lip's movements during pronunciation of the corpus by mathematical expressions called visual speech signals (global standardization).
- Preserving the bimodality and representing the relations between the corpus and lip's movement.
- Categorising the visual speech signals in a particular lip movement as visual word
- Increasing the ability of identification of lip's movements by defining visual word barcode, digital visual word, and 2D/3D signatures representations.
- Relating the feature points on the lip and defining single template of lips which is led to volumetric representations of the visual words.

In the field of visual speech recognition and animation, these approaches can introduce new path of analysing and synthesising the lip's movements.

1.6. POTENTIAL APPLICATIONS OF THIS WORK

The lip's movement in mathematical domain has vast variety of applications. The articulated words are classified as speech signal. Such signals could be used for enhancing speech signals perception in AVASR systems as well as animation and constructing a stand-alone system for synthesizing the visual speech. However, constructing such systems is out of the scope of this work. The practical implementation of the suggested methodologies in the AVSR systems and visual speech synthesisers will be considered as the feature work of this study.

1.7. OVERVIEW (ORGANIZATION OF CHAPTERS)

This thesis is organized as follows:

In Chapter 2, the backgrounds and literature review is presented and studied. It includes basic definitions in the field of audio-visual automatic speech recognition and as the main application of the current study on mathematical models of visual speech. Furthermore, fundamental methods and approaches are studied as polynomial

interpolation, geometric modelling, discrete-time representation of visual feature samples, pulse with modulation and source coding.

In Chapter 3, a method of designing the transcribed words from a transcribed audio-visual database will be suggested. In this chapter, the hierarchal analysis of the phonemes sequences base on their statistical information is addressed. The analysis leads to find a group of words with three phonemes from the beginning of word. This group of words will be the subject of visual speech analysis and visual speech signal synthesis.

The property of the region of interest will be defined in Chapter 4 which is the source of visual. The methods of extracting visual information from two non-native English speakers will be followed by selecting the appropriate video files will be represented. A comparison between automatic and manual visual speech sample extraction will be demonstrated. In order to obtain accurate visual speech sample sets, the extraction of visual speech samples will be performed manually.

In Chapter 5, processing of acquired data will be applied to the visual speech sample sets. Two types of data modifications are used to modify and organize the extracted visual information. The statistical analysis of the visual speech sample sets is determined the density function type of each visual speech sample sets before and after modification. Afterward, non-parametric density estimation used the visual speech sample sets density functions before and after modifications. The non-parametric approximation will be used in digital representation of visual words (Chapter 7), in order to select the optimum partitioning.

In Chapter 6, the main mathematical approach for deriving the visual speech signals is evaluated. Such evaluations are conducted by alternating the number of samples and their amplitudes, in term of interpolation error.

In Chapter 7, the main core of this thesis will be introduced for deriving visual speech signals and their signatures. In addition, the reconstruction performance of Barycentric Lagrange Interpolation is evaluated. The methods will allocate more identification properties and visual signatures to lip movement namely visual word barcode, digital visual word and 2D/3D signatures. The other type of 3D

representation is obtained by the volumetric modelling of the visual words. In the end, an alternative visual speech signal derived by sinc function will be compared to the visual speech signals obtained by BLI approach.

Discussion and conclusion about the suggested method and its outcomes along with the potential future work of this thesis for employing in the AVSR systems and visual speech synthesisers are presented in Chapter 8.

In Appendix, a general review of visual speech synthesising methods (Appendix A), log-likelihood tables (Appendix B), viseme grouping according to other researchers (Appendix C), the two main algorithms used in this work (Appendix D) and the mathematical expressions (Appendix E) are included.

Chapter 2

BACKGROUNDS

The words are forms of human thoughts that emerged as sounds and made by shaping articulatory organs. The articulatory organs are grouped into vocal tract and nasal tract. The vocal tract consists of the mouth (oral cavity) and pharynx while the nasal tract starts at the velum and stops at the nostrils (Rabiner & Jung, 1993). The most visually observable parts of articulatory organs in face-to-face communication are the lips.

Learning a language starts a few months after birth. In very basic level, toddlers react to the surrounding voices. Parents start to teach important words to their babies like 'Mum', 'Dad', baby's name, etc. Looking at that fundamental process of learning the language by babies attempt to repeat the words for realizing and memorizing them spoken by parents, reveals the physical pronunciation practicing. The visual information exhibited by parents, has equal or somehow more emphasis in comparison to words sounds. What has been observed by baby's eye is used to train the articulatory organs. Babies can understand the particular meaning of some words and react proportionally after such trainings. Then parents add in more complexity to their lessons. In parallel, the baby enhances its visual perception and generation abilities. In this way, the babies learn enough for verbal (mouth to mouth) communication. However, in deaf and mute cases this hypothesis is not applicable. Inspired by such interactive method of learning, in the human-machine interaction, same procedure could be employed.

2.1. PHONEME

Before studying the concept of phoneme, it is necessary to be familiarised with other basic terms. In this section, three important concepts of spoken language are described briefly as phone, phonology, phonetics that would describe the backbone of phoneme concept.

Any spoken language has been made up of sounds. In English language, an element of language is called phone. More specifically, phone is 'an elementary sound of spoken language; a simple vowel or consonant.' (Oxford, 1993).

In this section, the etymologies used in this work as phonology, phonetic, phoneme are reviewed. The meaningful combination of phones is examined in the field of phonology. Generating the most appropriate speech sound in a language has been studied by phonetics symbols. In other words, transcribing into text, the sound which make up words are phonetics. The international phonetic association suggested the standardized phonetic alphabet. It is denoted according to International Phonetic Alphabet (IPA) (International Phonetic Association, 2005). The IPA defines the texted symbols used to define each sound. In IPA each phone is represented by a symbol as shown in Table 2-1 and Table 2-2 where a word is represented by phonetic symbols in brackets. The phonetic symbols reduce the speaker's confusion about pronunciation. For example, in English language the word 'chaos' has been pronounced according to the IPA as [keɪɒs].

Phoneme is the basic distinguishable unit of the speech sound. Phonemes are the result of speech sound segmentation where if a segment (phoneme) is replaced by other, the meaning of phoneme string (speech sound) changes. In the literatures, phoneme is transcribed by one or two small English alphabet called phonemic symbols.

Generating the speech sound by test subjects closed to ideal pronunciation by natives of the language is addressed by phonetic alphabets. On the other hand, the most elementary and distinguishable unit of generated speech is categorised as phoneme. The phonemes are appearing in a meaningful sequential trend. This arrangement categorises is studied in the field of phonology.

2.2. CONSONANTS

The phonetic alphabet is represented and organized by the International Phonetic Association as it represented for consonants in Table 2-1. This table consists of 57 phonetic symbols. As it can be observed, instead of suggesting transcribed word examples for correct pronunciation, the places of articulation are used i.e. bilabial, labiodentals, dental, etc. as well as the manner of articulation which is related to directing of air flow to glottal channel, nasal and vocal tract i.e. plosive, nasal, trill,

etc. According to the Cambridge Advanced Learner's Dictionary (Cambridge, 2003), a consonants is one of the speech sounds or letters of the alphabet, which is not a vowel. Consonants are pronounced by stopping the air from flowing freely through the mouth, especially by closing the lips or touching the teeth with the tongue (Cambridge, 2003)'.

Table 2-1: The phonetic alphabets and their place of articulation revised 2005(International Phonetic Association, 2005)

CONSONANTS (PULMONIC) © 2005 IPA

	Bilabial	Labiodental	Dental	Alveolar	Postalveolar	Retroflex	Palatal	Velar	Uvular	Pharyngeal	Glottal
Plosive	p b			t d		ʈ ɖ	c ɟ	k g	q ɢ		ʔ
Nasal	m	ɱ		n		ɳ	ɲ	ŋ	ɴ		
Trill	ʙ			r					ʀ		
Tap or Flap		ⱱ		ɾ		ɽ					
Fricative	ɸ β	f v	θ ð	s z	ʃ ʒ	ʂ ʐ	ç ʝ	x ɣ	χ ʁ	ħ ʕ	h ɦ
Lateral fricative				ɬ ɮ							
Approximant		ʋ		ɹ		ɻ	j	ɰ			
Lateral approximant				l		ɭ	ʎ	ʟ			

Where symbols appear in pairs, the one to the right represents a voiced consonant. Shaded areas denote articulations judged impossible.

In Figure 2-1, the places of articulation in a human articulatory organs model are shown. For instance, the consonant [m] is produced by blocking the plosive kind of air behind the upper and lower lips which is called bilabial consonant.

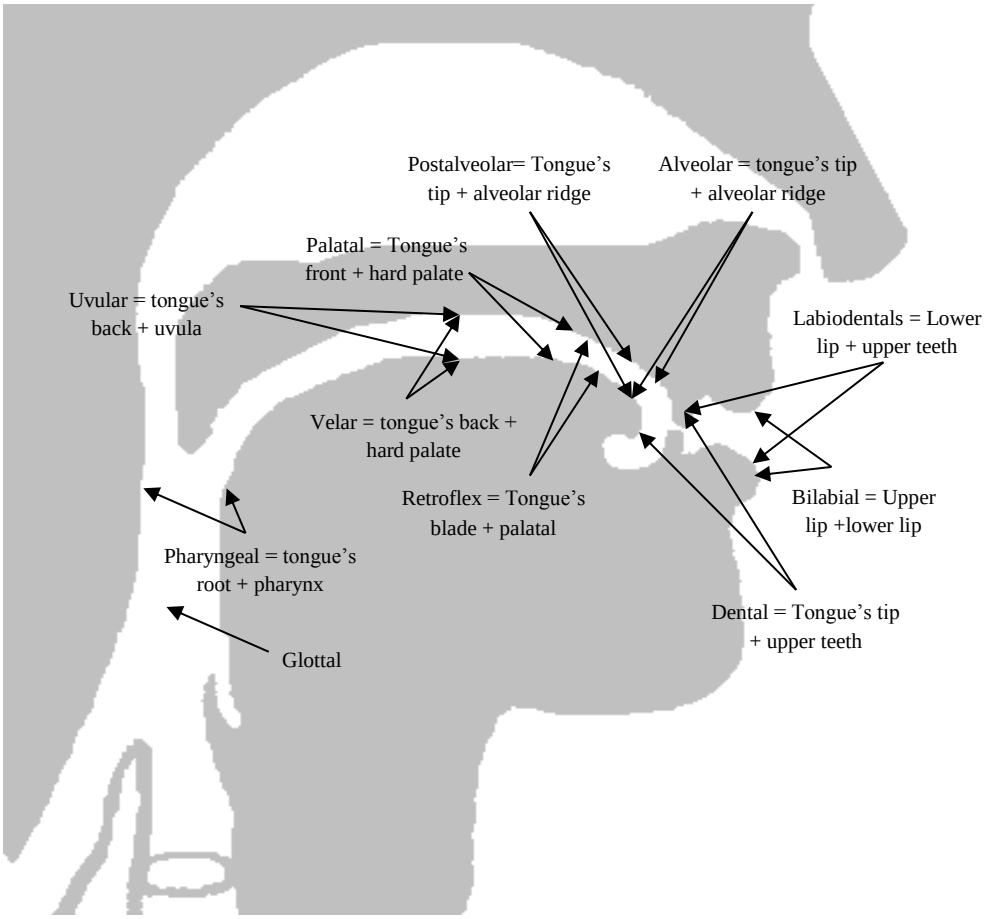


Figure 2-1: The places of articulation

The most visible places of articulating consonants are the labiodentals (lower lip and upper teeth), bilabial (upper lip and lower lip), and dental. Therefore, these places carry the most information for lip-readers and hence visual speech recognizers concentrate on these areas for the visual speech animation.

2.3. VOWEL

The second category of phonemes is vowels. These linguistic components are visually more distinguishable and visible than consonants. This effect is because of the shaping of the lips in absence of either the teeth or tongue to control the air flow.

Table 2-2: The place of articulation of vowels (International Phonetic Association, 2005)

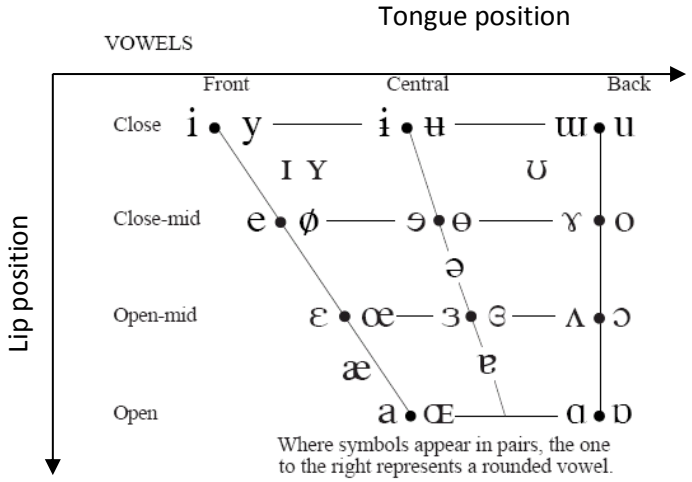


Table 2-2, shows the position of the lips and the tongue for different vowels. Vertical axis is the lip position and the horizontal axis is the tongue position.

2.4. VISEME

Defining a framework to relate speech units (phonemes) with their corresponding visual lip's shapes goes back as far as 1968, where the concept of visual phoneme (viseme) as visual representation of one or two phoneme was introduced by (Fisher, 1968).

2.4.1. VISEME CLASSIFICATION

The classification of visual speech is evaluated by the human perception of mouth movement (Williams & Katsaggelos, 2002). Such perception is conducted by lip-readers, hearing impaired and normal people. The resulting corpus contains variety of different phoneme combinations. As an observation, the combinations of sounds did not necessarily make meaningful speech sounds.

To ensure the test subjects concentrated on the position and measurement of the mouth, the tests are performed in the absence of sound. Subsequent analysis of the perceived movement of lip by the test subjects and the actual sounds resulted in what is called the 'confusion matrix'. In such matrix the 'correct' answer is the maximum vote from the observers. The measurement of the confusion matrix has been the topic of several studies (Williams, et al., 1997) and there was and still is a lack of agreement on what sounds to include and how to notate visemes.

2.4.2. PHONEME-TO-VISEME MAPPING TABLE

Finding the relationship between phonemes and visemes is a topic in visual representation of speech. Phonemes are usually grouped into viseme classes (Table 8-6) but there is no standard classification. For instance, the viseme group /p, b, m/, which was also detected by most of the researchers, indicates the visual similarities in observing the /p/, /b/ and /m/ consonants by lip-readers more than one phoneme.
Beside the visual similarity of different phonemes there are articulation differences between speakers (Rogozan, 1999) that reduce the reliability of visemes. Furthermore, each viseme is treated separately. In reality, there is transition between each phoneme (Co-articulation) which can affect the perception.

2.5. CO-ARTICULATION

In this section, the Co-articulation effect and its visual counterpart are studied in more detail. In a sequence of phonemes, a phone is affected by neighbouring phones. This effect in the speech domain is called Co-articulation. More specifically, the effect of proceeding sound refers to backward Co-articulation while the effect of upcoming phonemes is referred to forward Co-articulation (Cohen & Massaro, 1993). The forward Co-articulation appears by affecting the upcoming vowels. For instance, in the first consonants of words ‘dad’ and ‘door’ is /d/ which is affected by proceeding vowels /a/ and /o/. The backward Co-articulation occurs on the consonant /d/ in last part of words ‘bad’. The sound associated with the consonant /d / is affected by previous vowels /æ/ and /e/. Obviously, such effects also occur in visual domain. The Co-articulation effect results in a difference lip movements when uttering a single isolated phone in comparison with a sequence of phones. This has been studied by (Benguerel & Pichora-Fuller, 1982). Modelling the lip transition between visual phonemes was suggested by (Montgomery, 1980) who used a non-linear interpolation method. Another model of Co-articulation has been suggested by (Cohen & Massaro, 1993). This model allows the smooth transition between mouth shapes related to speech segments by defining dominance and blending functions. Another solution of transition between visemes was addressed by the Hermite interpolation (De Martino, et al., 2006).

Appearing the confusion and Co-articulation effects are not included in viseme concept. These facts imply that the viseme concept may not be the appropriate building block of visual speech expression. If visual recognition unit does not cover such criteria, it will not be a reliable interpreter for visual phonemes. In this work, BLI is suggested as mathematical framework for standardizing the visual phoneme by categorizing the phonemic structures of words.

2.6. LIP READING

Lip reading by human has been mastered for a long time. Limited number of people who called lip-readers learnt to use and interpret visual information. A correlation between generated sound and the lip formation makes the lip-readers to estimate the possible phones combination. Lip reading allows comprehension of speech by observing the lip movement. Not so long ago this interaction was restricted to human face to face communication in special situations. Today the communication concept has advanced beyond human society and is a key element in human-computer inters action. These topics are explained in the following sections.

2.6.1. LIP READING BY HUMAN

The act of understanding the spoken word by 'reading' the movement of lips is generally referred to as lip reading. Lip readers 'read' the visual clues and combinations along with a grammatical correlation of the language to form a coherent comprehension of the spoken word. Observing the lip movement is not sufficient to fully comprehend the spoken word since other face components e.g. eyes, cheek, and eyebrows transfer additional information.

2.6.2. VISUAL SPEECH RECOGNITION

Studying of lip reading makes it possible to extend the concept to human–machine communication. A simplified process of visual clue perception by human can be applied at the machine level using mathematical algorithms and models. Machine perception of speech is initially limited to analysis of the audio features often referred to as automatic speech recognition (ASR). Although highly successful, ASR does have a major challenge when the speech signal is corrupted with noise. To address

this problem, the visual clues can be used (Sumby & Pollack, 1954) and (McGurk & MacDonald, 1976).

The importance of visual clues to human understanding of the spoken word was demonstrated by McGurk effect (McGurk & MacDonald, 1976). The McGurk effect states when a person receives conflicting audio and visual clues. For example, when a person heard the sound /ba/ but watches the lip movement of /ga/, the perception of the subject is neither /ga/ nor /ba/ but /da/.

Generally, there are five methods of interpretation of the audio-visual information (Deligne, et al., 2002):

- Audio-visual speech recognition (AVSR)
- Audio-visual automatic speech recognition (AVASR)
- Visual-only automatic speech recognition (speech-reading)
- Visual speech animation
- Automatic speech recognition enhancement systems

In the mid 1980's the first automatic AVSR was introduced and used the visual modality to enhance the speech recognition (Petajan, 1984). In speech-reading systems, the visual speech components are extracted, analysed and a stream of symbols are outputted (Potamianos, et al., 2004). In AVASR, the visual speech information components are analysed in parallel with the audio speech channel and are fused to deliver a more accurate representation of the spoken word (see Figure 2-2).

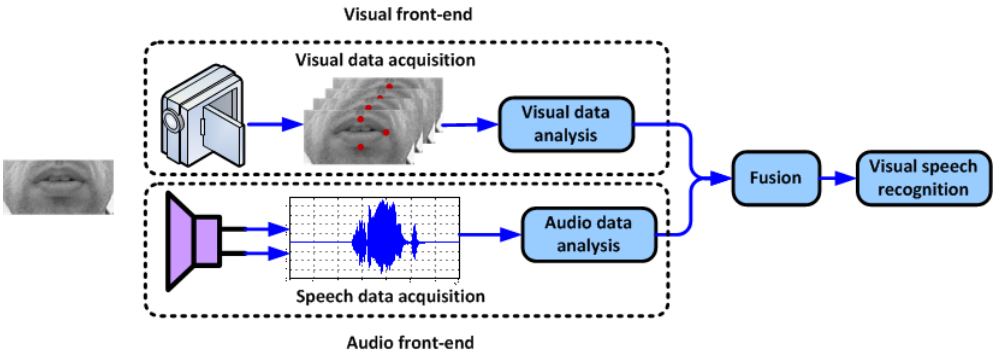


Figure 2-2: AVSR system

The visual front end in an ASR aims to extract the visual features from a video signal in three steps:

a) Face detection
b) Mouth localization
c) Lip tracking algorithms

Localizing the face is done by using image-processing methods based on the pixel values. This is usually performed by classification or by edge detection between the face and the background (Petajan, 1984). The other approaches are snakes (Kass, et al., 1987), deformable template of mouth (Rao & Mersereau, 1995) and cubic polynomials (Seguier & Cladel, 2003), (Stillittano, et al., 2009) and (Eveno, et al., 2002). More sophisticated method use Active Appearance Model (AAM) (Cootes, et al., 1998) for face localization (Edwards, et al., 1998).

After face detection and mouth localization, the lip region needs to be found. This process is referred to as locating the region-of-interest (ROI). One common method is based on the statistical models and it is called Active Shape Model (ASM) (Cootes, et al., 1995) and (Luettin. & Thacker, 1997).

After localizing the mouth region, in order to generate a suitable the stream of visual information lip tracking algorithms are employed. Among the tracking methods, snake (active contour models) (Kass, et al., 1987), AAM, deformable template of mouth and ASM are widely used.

After the pre-processing phase, which includes implementation of face detection, mouth localization, and lip tracking algorithms, the lips visual features need to be extracted. In other words, it was desirable to select a set of samples or values on detected mouth which could represent the deformation of lips efficiently. Extracting the lip's parameters in time or frame intervals mainly have been performed by using image processing on the texture or electronic sensors.

The visual features of lip can be represented by three different categories (Potamianos, et al., 2004) based on the possibilities of identifying the lip in image:

a) Appearance based features
b) Shape based features
c) Combination of appearance based features and shape based features

In the first category, the pixel colour value is used to extract information about the lip's shape (also called appearance based features) while in the second approach, the

spatial value of pixel was the subject of consideration (shape based features). The third approach is a combination of both first and second approaches.

In the appearance based approach, the pixels in the ROI is reduced from two dimensions to one by pixel concatenation. Information relating to the lips shape is directly contained within the one dimensional pixel array. This technique has been used by (Bregler & Konig, 1994) and (Rogozan & Deléglise, 1998) or colour pixels (Chiou & Hwang, 1997). Other transformation approaches have also been employed as Principle Component Analysis (PCA), Discrete Cosine Transform (DCT) and Linear Discriminant Analysis (LDA).

Shape based lip identification uses mathematical models to determine the shape of lips. A Model of mouth containing of 130 vectors was developed by (Montgomery, 1980) whereas trigonometric functions was introduced by (Yuille, et al., 1992). The authors in (Erber, et al., 1979) used the Lissajou's parametric equation to figure the mouth shape (Cohen & Massaro, 1993).

In the combined approach, the appearance based algorithm are used to identify the lips followed by the extraction of movements, which describes the shape of lip.

Another strategy for visual feature allocation to lip based on combination of appearance-based features and shape based features proposed in (Potamianos & Graf, 1998) where the inner and outer lip's contour estimation was adopted from (Graf, et al., 1997) similar to the method used by (Chan, et al., 1998). The corresponding lip parametreasation is depicted in Figure 2-3 (a). Unfortunately, the procedure used in these methods does not uniquely define the shape of the lips in a coherent way. To include information about the shape of the lips, (Teissier, et al., 1999) added the inner surface of the mouth as an additional parameter where it demonstrated in Figure 2-3 (b). Again, the coherency between parameters does not exist.

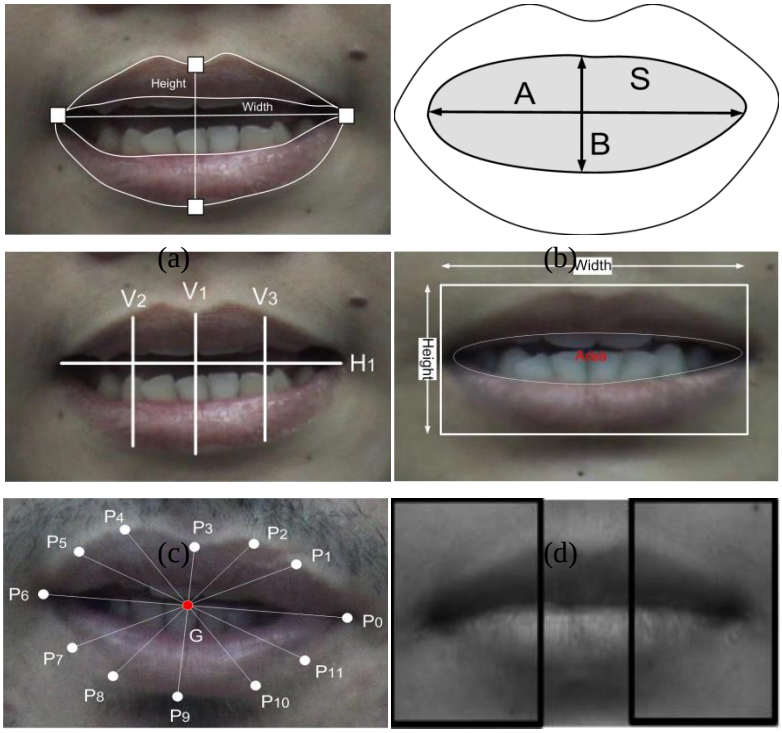


Figure 2-3: Geometrical representation of lip used by (a) (Graf, et al., 1997), (b) (Teissier, et al., 1999), (c) (Silveira, et al., 2003), (d) (Lyons, et al., 2004), (e) (Saitoh & Konishi, 2005), (f) (Zhao, et al. 2009)

Another method has been suggested by (Silveira, et al., 2003) who parameterized the mouth by four distance features V_1, V_2, V_3 and H_1 as it shown in Figure 2-3 (c). H_1 measures the lip's dimension horizontally by connecting the corner points of the mouth while V_1,V_2 and V_3 measure the vertical dimension of lip. The V_1 located on the middle of H_1, V_2 on the middle of lip's left corner point and V_1 and finally V_3 on the middle of lip's right corner point and V_1. The vectors V_2 and V_3 do not carry much information since generally their distances to V_1 are almost equal. A similar approach to (Teissier, et al., 1999) was undertaken by (Lyons, et al., 2004) wherethe lip's aspect ratio (ratio of height and width) and the extracted mouth cavity area are chosen as feature representation as it depicted in Figure 2-3 (d). The lip features in (Saitoh & Konishi, 2005) contains of 12 distances measured from the centre of shown in Figure 2-3 (e). However, computational complexity over the previous research method was increased. Extracting visual features using the appearance based approach in (Zhao, et al., 2009) was depicted in Figure 2-3 (f). The lip area was divided into two rectangles which their pixel values would be statistically analysed. The major disadvantage of this approach was complexity of processing images before analysing the lip shapes.

This method of lip segmentation tries to reduce the computational cost of mouth analysis by using only corner sides of lip region and neglecting the middle region.

In addition, the visual feature points are extracted by an infrared sensor, which detects an emitting diode placed on the midpoint of lower lip (Lucero, 2002). The other method for defining the visual feature was employing optical flow (Mase & Pentland, 1991). The optical flow algorithm shows the velocity field in an image. Therefore, the movement of lip can be interpreted velocity vectors. One of the main advantages of using optical flow algorithm was related to a covering both lip localization and tracking. In visual speech synthesis, the visual phonemes provided by the lip reading systems are combined to animate a 2D or 3D lip. An overview of visual speech synthesis is represented in the Appendix A.

In this work, in order to achieve the aims of the research, the structure of lips are studied without using face detection, localization, or lip tracking algorithms. The structure of lips as visual features was the subject of visual speech analysis and will be used for visual speech signature synthesis.

2.7. VISUAL SPEECH DATA PROCESSING

In AVASR the aim was to generate or enhance the audio speech signal from visual speech signal while in visual speech synthesis, generating the visual speech signal from audio speech signal was desirable. In most literatures, the statistical analysis of visual modalities was used for speech reading or visual speech animation. Using the neural network as the classifier, the visual speech data is represented to the input and output of the network, which was called training phase, according to application (visual speech generation or perception). For instance, the neural network classification applied to AVSR (Lewis & Parke, 1987). Afterward, a call-back algorithm trains the network weights.

The other widely method used in visual speech recognition and visual speech synthesis system for defining the dynamics of lips geometries was called Hidden Markov Model. The HMM is a stochastic model for time-series data, which is a statistical tool for representing the probable events. In this approach, events are modelled by states and the sequential appearance of them can be generated via probabilistic transitions. The Markov process (Markov, 1913) was the foundation of HMM (Baum, et al., 1970). In the Markov process, each state emits a single event or

observation. The observations are probabilistic functions of states. The concept does not include the process of generating sequence of observation. However, it suggests a Markov model to explain the sequence of observations. The probability of observation sequences was simply the product of transition probabilities. In most cases, the states are emitting events in common. Therefore, exact determination of states collaborating in observation generation was difficult. In other words, the state sequence was hidden. The HMM model parameters are (Rabiner, 1989):

$$\lambda = (A, B, \pi) \tag{2-1}$$

where A, B and π are state alphabet, state observation set and initial probability array, respectively. After creating model parameters in Eq. (2-1) by evaluating HMM, it is possible to determine the probability of observing a sequence:

$$O = O_1, O_2, O_3, \dots, O_T \tag{2-2}$$

where T is the number of observations. Modelling a process by HMM needs examples of the process that was called learning or training the HMM. The learning process has two approaches as supervised learning and unsupervised learning. If the desirable input and output are allocated to the HMM, the process will be called supervised training while in the unsupervised training only input was associated to the HMM.

In AVSR systems, the HMM model was used by speech reading system proposed by (Goldschen, et al., 1994) which was applied to the lip-reading system introduced (Petajan, 1984). The model also used by (Williams, et al., 1997), (Xie & Liu, 2006) and (Ling & Wang, 2006) where the speech feature parameters are used for tainting the model. Then when the trained HMM was represented by the speech signals, the lip animation was achieved. In the proposed model by (Potamianos, et al., 1998), (Rabiner & Juang, 2008) the parameters of visual features related to the phonemes sequences trained the model. When the optimum probable states are defined, the system can be used for speech recognition.

A 3D lip model was fitted to lips image (Revéret & Benoît, 1998) and the amount of similarities between the feature point on the model and actual lip was measured. In Figure 2-4, the results of such measurement by tracking the lip feature point's distances are illustrated where the circles are representing the tracked 3D model and crosses are the marking the actual lip movement on each frame. Their results did not conclude the dynamic of articulating lip with a mathematical expression.

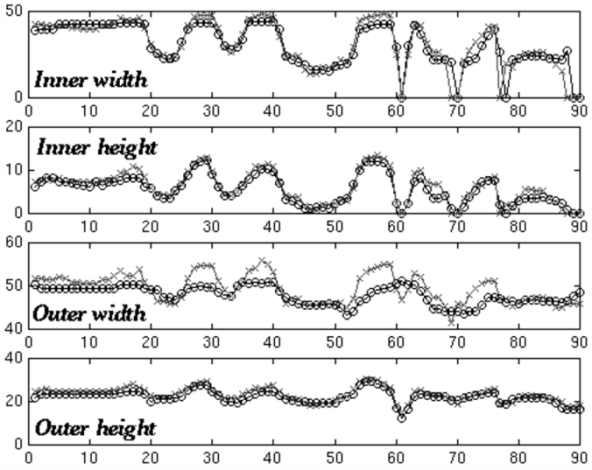


Figure 2-4: The geometric feature extraction by tracking the lip's model marked with circles and the actual lip marked with crosses (Revéret & Benoît, 1998)

In the suggested lip-reading system by (Potamianos, et al., 1998), the visual features related to the width, height, and area of lip are tracked over frame/time domain during articulating a sequence of digits '81926' as shown in Figure 2-5. Such trajectories of visual features are used for training a HMM model and had not been mathematically formulated.

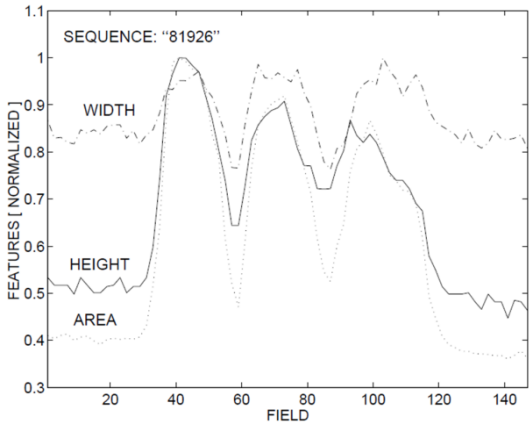


Figure 2-5: The three geometric features from lip tracked over articulating the sequence '81926' (Potamianos, et al., 1998)

The acquired data (from movement of lip) in (Lucero, 2002) was formulated by B-spline cure with the order six. The main disadvantage of using B-spline for formulating the lip's movement was the lack of defining a single expression for the sample sets rather on segmented intervals.

The extraction of feature points using the appearance-based approach in (Zhao, et al., 2009) was suggested a statistical visual features modelling by concatenating the

image sequence portions in block volume, which was depicted in Figure 2-6 (a). Using an image processing approach called Local Binary Pattern (LBP), the lip texture between successive images in grey scale are transformed to binary codes.

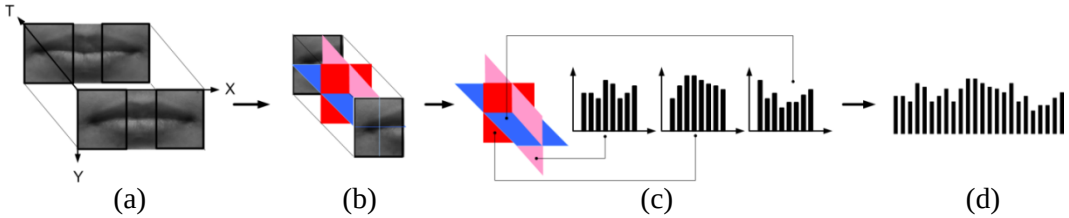


Figure 2-6: The process of visual feature extraction used by (Zhao, et al., 2009)

Afterward the LBP applied to the other two orthogonal plans, which codes the spatial and temporal axis as shown in Figure 2-6 (b). The histogram of these binary orthogonal planes was calculated and represents the visual feature. By concatenating the histograms, two important characteristics of the appearance of the lips and their motion are extracted, which was obtained in Figure 2-6 (c). The sequences of histograms are joined in Figure 2-6 (d). In other words, their method not only extracts the visual feature but also express the deformation of visual features dynamically.

The method suggested by (Birkholz, et al., 2011) used a time-variant dynamic system used for modelling the articulatory relate facial parts as tongue tip, tongue blade tongue, jaw, dorsum, upper and lower lip. In Figure 2-7 (a), the results for articulating the [CVCVCVCV] sequence with normal speaking rate is shown. In Figure 2-7 (b) the slower speaking rate is shown. This method is computationally intensive.

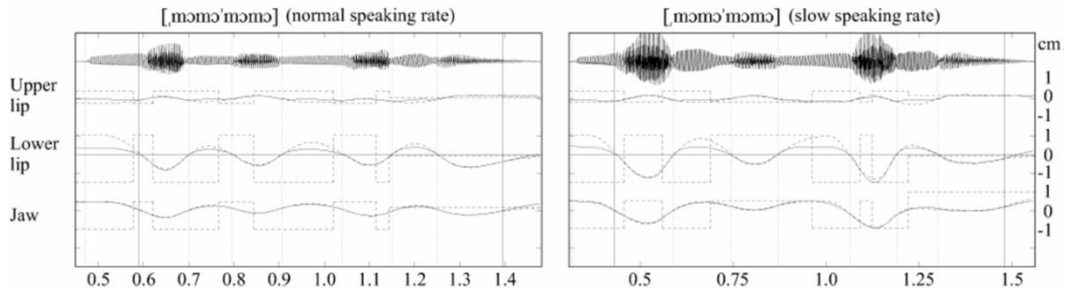


Figure 2-7: The results of tracking the upper lip, lower lip and jaw in articulating the [CVCVCVCV] sequence with normal (a) and slow (b) speaking rates (Birkholz, et al., 2011)

2.8. AUDIO-VISUAL DATABASE

Audio-visual database is a collection of speech signal files along with the corresponding video files and transcription of the text corpus. The text corpus of the

audio-visual databases, which was adopted for this work, was one of the important components. It acts as feeding materials to the system in form of phones, words, or sentences. Most of the databases are providing the text corpora in form of dictionaries with phonemic transcriptions.

One of the most well-known sources of ASR, which provides varieties of the audio-only databases bodies, is the Linguistic Data Consortium (LDC). Among the databases, the TIMIT Acoustic-Phonetic Continues Speech Corpus was chosen. This database was widely used in the ASR systems like Hidden Markov Model Toolkit (HTK). The TIMIT has a corpus of audio-phonemic information of speech. The data was gathered from 630 speakers in 8 dialects of American English; each was reading up to 10 sentences. This database was a collaborating work of the Massachusetts Institute of Technology (MIT), Texas instruments, Inc. (TI) and Stanford Research Inc. (SRI).

The phonemic symbols and their examples used in the TIMIT's transcribed dictionary have been represented previously by (Garofolo, et al., 1993). Table 2-3 provides more details about the manner of articulation and examples. Table 2-3 consists of 32 phonemic symbols for consonants grouped by their manners of articulations and 20 phonemic representations of vowels.

Table 2-3: The phonemic symbols in the TIMIT Acoustic Continuous Speech Corpus (Garofolo, et al., 1993)

	Symbol	Example		Symbol	Example
stops	b	bee	Vowels	iy	beet
	d	day		ih	bite
	g	gay		eh	bet
	p	pea		ey	bait
	t	tea		ae	bat
	k	key		aa	bott
	dx	muddy, dirty		aw	bout
	q	bat		ay	bite
Affricates	jh	joke		ah	but
	ch	choke		ao	bought
Fricatives	s	sea		oy	boy
	sh	she		ow	boat
	z	zone		uh	book
	zh	azure		uw	boot
	f	fin		ux	toot
	th	thin		er	bird
	v	van		ax	about
	dh	then		ix	debit
Nasals	m	mom		axr	butter
	n	noon		ax-h	suspect
	ng	sing			
	em	botton			
	en	button			
	eng	washington			
	nx	winner			
Semivowels and Glides	l	lay			
	r	ray			
	w	way			
	y	yacht			
	hh	hay			
	hv	ahead			
	el	bottle			

In the TIMIT's dictionary the phonemic sequence of each word was represented between two slashes (/). For example, the word 'symbolic' phonetic notation is [sɪmbɑlk] while it's phonemically represented as /s ih m b aa l ih k/. One of the main features of this dictionary is the usage of space in phonemic sequence. This characteristic provides a great benefit to this work where analysing the phonemic sequence are the main approach for unifying and selecting a group of words. The procedure is described in Chapter 3.

The arrangements of phonemes in transcribed form are considered as one of the important subjects in visual speech analysers and synthesisers. The transcribed words are called corpus or simply dictionary, which could form meaningful or non-meaningful words. The concept presented in this thesis uses such corpora for development of a novel lip-reading system. There are several different corpora presented in the literatures and differ in the structure of phonemes for generating lip

movements. The structures of the corpora are shown in Table 2-4 where C and V refer to consonant and vowels phonemes respectively. Each set corresponds to a specific combination of isolate phonemes.

In the first category of transcribed data, the isolated phoneme was used for lip reading (Petajan, 1984) or animation (Bregler, et al., 1997). In the second scenario, a specific arrangement of vowels or consonants was fixed for visual analysis of speech. A corpora was suggested by (Montgomery, 1980) which used the /CVCVC/ synthesisers while (Adjoudani & Benoît, 1996) used /$V_1CV_2CV_1$/ corpus. The random combination of phonemes such as /aCa/ has also appeared in the corpora structure by (Su & Silsbee, 1996). The phoneme arrangements as a part of a French sentence 'C'est pas /VCVCVz/?' was examined by (Revéret & Benoît, 1998). (Czup, 2000) used triphone sequences /V_1CV_1/ and /C_1VC_1/ corpora. For a Brazilian Portuguese facial animation system proposed by (De Martino, et al., 2006), the corpora consisted of two types of phonemes combinations as /CV_1CV_2/ and /V_1V_2/ while the /VCV/ structure was used by (Cox, et al., 2008). This model was used for training a HMM in ASR system with different four conditions under seven levels of noise.

Table 2-4: The structure of phonemes used as corpora

Corpora	Authors	Language
/CVCVC/	(Montgomery, 1980)	English
/$V_1CV_2CV_1$/	(Adjoudani & Benoît, 1996)	English
/aCa/	(Su & Silsbee, 1996)	English
'C'est pas /VCVCVz/?'	(Revéret & Benoît, 1998)	French
/V_1CV_1/ and /C_1VC_1/	(Czup, 2000)	Hungarian
/CV_1CV_2/ and /V_1V_2/	(De Martino, et al., 2006)	Brazilian Portuguese
/VCV/	(Cox, et al., 2008)	English
/CVCV/	(Melenchón, et al., 2009)	Castilian Spanish
/CVCVCVCV/	(Birkholz, et al., 2011)	Not stated

The suggested structure of corpus of Castilian Spanish used by (Melenchón, et al., 2009) consisted of /CVCV/. Their reason for using such structure was supported by a strong statement, which claims more than 80% of Castilian Spanish words flow /CV/ structure (de Vega, et al., 1992). A suggested phoneme arrangements by (Birkholz, et al., 2011) produced /CVCVCVCV/ sequences. These methods are also noted in Table 2-4. Unfortunately, no evidence for choosing such arrangements of phoneme sequences was presented by the authors. The main disadvantage of such corpora arrangements is the limitation to the number of meaningful words.

2.9. GEOMETRIC MODELLING

In this section, details of parametric curve construction methods will be presented. which are called the Bezier curve, B-spline curve and Hermite curve. These methods are the main mathematical tools for modelling shapes and curves. The term construction has been used instead of interpolation since the data samples may not necessarily be located on the generated curve. The samples, from which curves are generated, are called control points. If the samples of a deterministic function are considered as control points, the generated curve will 'approximate' the function. On the other hand, if the control point's geometrical arrangements lead to a curve passing through (or fitted to) the sample set, it will be considered as a type of interpolation.

2.9.1. PARAMETRIC CURVE

The parametric representation of a curve is a method of defining the explicit equation of the curve $y = f(x)$ by two other functions $y = y(u)$ and $x = x(u)$ using a parameter u. The parameter u, which uniformly spans an interval $u = [a, b], a \neq b$ where a and $b \in \mathbb{Z}$, defines a continuous vector space. This interval is often normalized so that $u \in [0, 1]$. If a curve is represented by an explicit equation, the curve's range and domain can be analysed using the two parametric equations. A simple example of such representation is the unit circle given by equation:

$$x^2 + y^2 = 1 \tag{2-3}$$

The variables x and y in Eq. (2-3) can be formulated by two parametric functions as:

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} sin(u) \\ cos(u) \end{bmatrix} \tag{2-4}$$

2.9.2. BEZIER CURVE

A set of n control points defines the Bezier curve and determines its degree (Mortenson, 1997). The degree is defined by the basic functions, which are the fundamental components of the Bezier curve generation methodology. The main features Bezier curves are:

a) This family of curve modelling relates a set of at least 3 points in the space.

b) The degree of curve depends on the number of control points.
c) The basic functions are Bernstein polynomials.
d) The approximated curve is located inside a polygon that is made by control points while the starting and stopping control points are included to curve definition.
e) Changing the position of any points affects the whole curve and there is no any local control over curve's shape.
f) If the end points coincide then a closed curve is obtained.
g) If the control points p_1 and p_{n-1} are collinear, C^1 continuity (its first derivative) will be preserved otherwise the C^0 continuity was appeared in the end points.
h) The minimum number of vertices to make a closed curve, which does not preserve C^1 continuity, is four.
i) The basic functions are defined over normalized domain $u \in [0, 1]$.

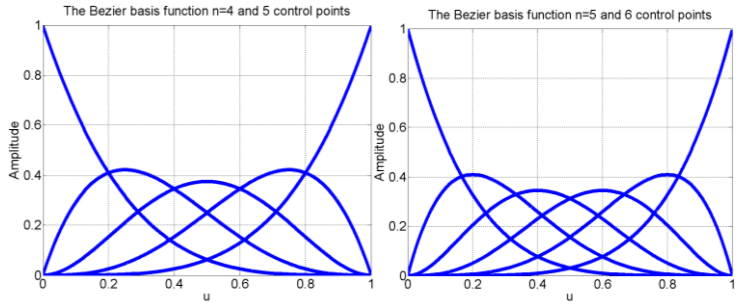
Since all the control pointes are used to generate the curve, a small displacement in the position of one of the control points will affect the shape of the curve throughout the region and not just in the location of the change. The Bezier curve equation is (Mortenson, 1997):

$$P_{Bezier}(u) = \sum_{i=0}^{n} p_i . B_{i,n}(u) \qquad 0 \leq u < 1 \tag{2-5}$$

The basic functions are defined as (Mortenson, 1997):

$$B_{i,n}(u) = \binom{n}{i} . u^i . (1-u)^{n-i} \tag{2-6}$$

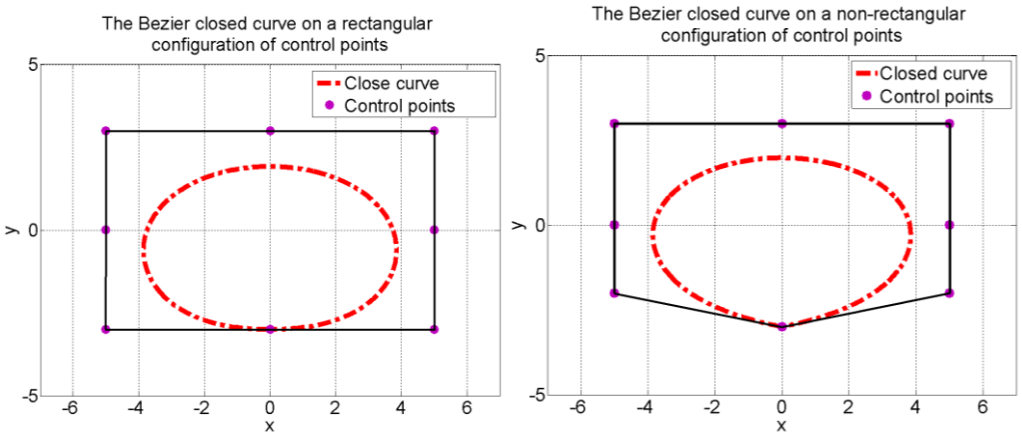
In Figure 2-8, two sets of basic functions are shown.



(a) (b)

Figure 2-8: Bezier (Bernstein) basic functions, (a) for $n = 4$ and five control points and (b) for $n = 5$ and six control points

In Figure 2-9 (a) a rectangular arrangement of control points with C^1 continuity while in Figure 2-9 (b) a convex arrangement of same points with C^0 continuity are shown. The minimum number of vertices to make a closed curve, which does not preserve C^1 continuity, is four.



(a) (b)

Figure 2-9: Closed Bezier curve for six points (a) with C^1 continuity, (b) same number of points with C^0 continuity

The main disadvantages of this family of curve modelling are:

a) The curve only passes through the end points.
b) The number of vertices controls the degree of curve causing an increase in the degree of curve.
c) Therefore, the Bezier curve is not a suitable candidate for modelling the lip movement.

2.9.3. B-SPLINE CURVE

In this section, modelling of curves using the B-spline curve is presented (Mortenson, 1997). Schoenberg, 1946, introduced the term of B-spline and its concept. The main features Bezier curves are:

a) It includes or excludes the end points (uniform and non-uniform B-spline curves)
b) Despite the Bezier method, the numbers of control points do not affect the degree of basic polynomial.
c) The B-spline domain is spanned by so-called knot values which are spaced uniformly or non-uniformly.
d) The knot vector can be normalised between zero and one
e) The curve will be uniform if the B-spline domain is uniformly spaced otherwise it is non-uniform.

The B-spline expression is (Mortenson, 1997):

$$P(u) = \sum_{i=0}^{n} p_i . N_{i,k}(u) \qquad 0 \leq u < 1 \qquad (2\text{-}7)$$

where p_i is the control point and $N_{i,k}(u)$ is the basic function. In Eq. (2-7), the resulted curve, which is weighted by $(n+1)$ control points, is summation of a set of basic functions. The degree of the basic functions is $(kn-1)$. For the non-uniform B-spline, the basic polynomials are defined by two expressions; when the order is one and the other for higher orders. This time the variable u is defined over different spans that are connected in knot space. The equations are:

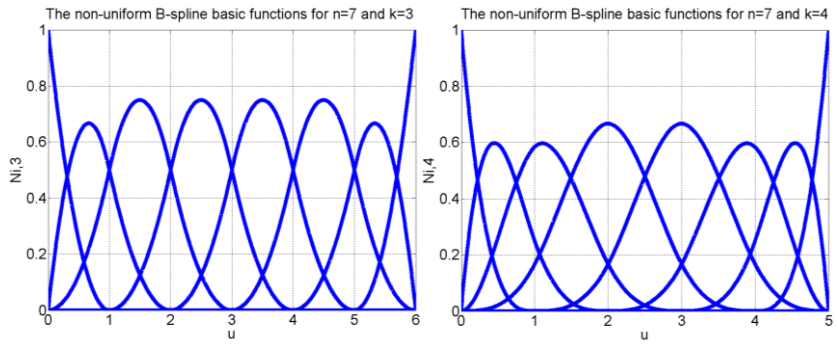
$$N_{i,1}(u) = \begin{cases} 1, & t_i \leq u < t_{i+1} \\ 0, & otherwise \end{cases} \qquad (2\text{-}8)$$

and:

$$N_{i,k}(u) = \frac{(u - t_i)N_{i,k-1}(u)}{t_{i+k-1} - t_i} + \frac{(t_{i+k} - u)N_{i+1,k-1}(u)}{t_{i+k} - t_{i+1}} \qquad k = \{2,3,4, \dots\} \qquad (2\text{-}9)$$

The t_i variable is knot value and a set of them combines to form the knot vector. The overall span, which has been divided into knot space, was u. Therefore, each basic polynomial can effect only on few number of knots.

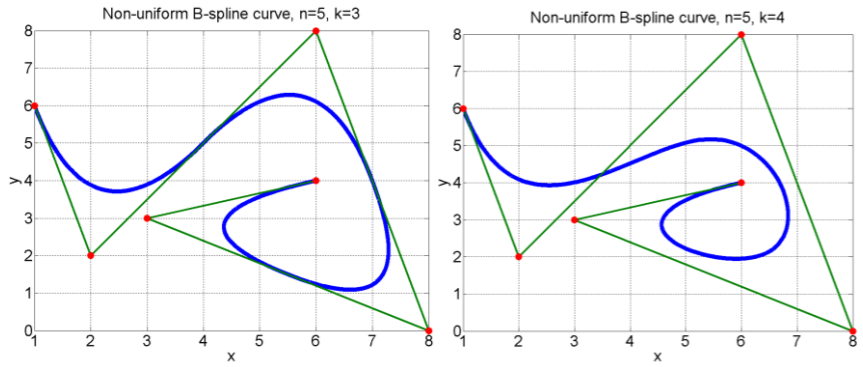
For instance, the basic polynomials defined by Eq. (2-9) for $n = 5$, $k = 3$ and $n = 5$, $k = 4$ are shown in Figure 2-10 (a) and (b) while Figure 2-10 (c) and (d) they are depicted for $n = 7$, $k = 3$ and $n = 7$, $k = 4$, respectivly.



(a) (b)

Figure 2-10: B-Spline Basic Functions for (a) $k = 3$ and $n = 7$, and (d) $k = 4$ and $n = 7$

One of the possible curves, which approximate a set of six points, was shown in Figure 2-11.



(a) (b)

Figure 2-11: The non-uniform (a) 2nd degree ($k = 3$) and (b) 3rd degree ($k = 4$) B-spline curves with 6 control points ($n = 5$)

The main disadvantage of this family of curve modelling is the multiple equations for a single curve. Therefore, each curve that defines by B-spline method has a set of equations and cannot be a suitable candidate for the aim of this thesis.

2.9.4. HERMITE CURVE

The cubic Hermite curve (Bartels, et al., 1998) can be used for interpolating a sample set. The main features of Hermite curve are:

a) The sample set should be piecewise.
b) The Hermite curve defines between two points by parameterization of the point's range and domain. The curvature has been defined by the tangent vectors on the first point and second point by substituting 0 and 1, respectively.
c) There are four basic functions, which define the Hermite curve defined as:

$$F_1(u) = 2u^3 - 3u^2 + 1 \tag{2-10}$$

$$F_2(u) = -2u^3 + 3u^2 \tag{2-11}$$

$$F_3(u) = u^3 - 2u^2 + u \tag{2-12}$$

$$F_4(u) = u^3 - u^2 \tag{2-13}$$

where $u \in [0, 1]$ and the Hermite curve is defined by:

$$P(u) = F_1(u)P(0) + F_2(u)P(1) + F_3(u)P'(0) + F_4(u)P'(1) \tag{2-14}$$

where $P'(.)$ was the first derivative of the cubic curve. Further simplification of Eq. (2-14) using subscripted notation of end points results in:

$$P(u) = F_1(u)P_0 + F_2(u)P_1 + F_3(u){P_0}^u + F_4(u){P_1}^u \tag{2-15}$$

2.10. CONTINUOUS-TIME AND DISCRETE-TIME SIGNALS

A signal is something that conveys information about a physical phenomenon during an observation period. Signals are defined on an independent variable that defines their domain and range. The domain of signal can be either continuous or discrete. The continuous-time signals are defined by continuous independent variable whilst discrete time signals by discrete independent variable (Oppenheim & Schafer, 1999). In term of statistical properties, a signal can be stationary (statistical properties do not

vary over time) or non-stationary (statistical properties vary over time). In addition, a signal can be pure or affected by noise.

2.11. DISCRETE-TIME REPRESENTATION OF VISUAL SPEECH SAMPLES

In this work, the discrete-time (DT) representation of signals is used to accommodate a convenient representation of extracted feature points of the lips. This method reduces the complexity of vector calculus in processing the extracted visual data by categorizing them as samples sets. Therefore, the visual speech samples are treated as values with no dependency on direction in the Cartesian space. The physical movement of articulating lip follows a continuous volumetric path. The lip motion parameterization allows complexity reduction and practical analysis of such volumetric information. Therefore, continuous lip motion can be extracted as indexes of lips geometry during articulation by selecting a limited number of feature points. As shown in Figure 2-12, the observation of such behaviour has been provided by a recording device and represented by values on discrete time spans in covering frame sequence.

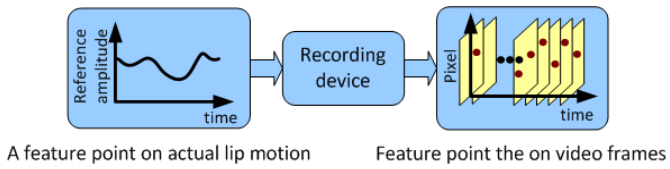


Figure 2-12: The actual relation between dynamic feature points and extracted feature points

Observing other signals dimension by the amplitude leads to classifying signals into analogue and digital. The analogue signal consists of unlimited values despite the limited range of digital signal. In addition, the digital signals are defined over discrete-time intervals. In this work, the input of the suggested system is digital signal (sequence of digits) while the outputs form continuous-time signals. The digital constructed signal appears in coding scheme of signal. The fundamental function, which has frequently been used to define continuous-time signal, is the Dirac delta function. The Dirac delta or unit impulse function $\delta(t)$ is an abstract impulse tap in $t = 0$ and defined in term of unit amplitude as follow (Oppenheim & Schafer, 1999):

$$\delta(t) = 0 \qquad t \neq 0 \tag{2-16}$$

$$\int_{-\infty}^{\infty} \delta(t) = 1 \tag{2-17}$$

The correspondence of unit impulse function in discrete-time domain is referred to as the Kronecker delta function or discrete-time unit sample function that illustrated in Figure 2-19 and defined as (Oppenheim & Schafer, 1999):

$$\delta[n] = \begin{cases} 1 & n = 0 \\ 0 & n \neq 0 \end{cases} \tag{2-18}$$

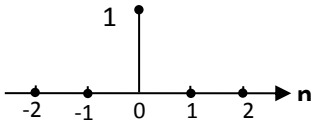


Figure 2-13: The unit impulse (Kronecker delta) function

For example for example, there is a band limited signal $x[n]$, which it members (samples) exist on discrete-time instances n, as shown in Eq. (2-19). Multiplying shifted version of Eq. (2-18) by k and scaling by corresponding sample leads to:

$$x[k]\delta[n-k] = \{a_{-1}\delta[n+1], a_0\delta[n], a_1\delta[n-1], a_2\delta[n-2]\} \tag{2-19}$$

A more compact relation between delta function and sample set $x[n]$ in Eq. (2-19) is (Oppenheim & Schafer, 1999):

$$x[n] = \sum_{k=0}^{2} x[k]\delta[n-k] \tag{2-20}$$

where k implies the sample values located on discrete-time axis n.

2.12. POLYNOMIAL INTERPOLATION

In order to find a mathematical tool for the explicit expression of the lip's movement, the interpolation approach was used. The interpolation method allocates a polynomial function to a set of data points by adding extra samples in between sample pairs. The basic idea of interpolation theory is the amplitude of intermodal points is determined by the application of a polynomial function whose degree is the one less that the number of nodes. This can be represented mathematically as $P(x) = \sum_{i=0}^{N} \alpha_i x^i$. The

values of α are chosen so that the polynomial goes though the given data points. The basic functions are exponential function and $i \in \mathbb{Z}$. This concept can be generalised by noting that the power sequence $\{1, x, x^2, \cdots, x^N\}$ forms a set of basis functions and these can be replaced by other functions sets $\varphi_i = \{\varphi_o, \varphi_1, \varphi_2, \varphi_3, \ldots \varphi_N\}$ (Mertins, 1999). This combination can be stated as (Mertins, 1999):

$$P = \sum_{i=0}^{N} a_i \, \varphi_i \tag{2-21}$$

(Schoenberg, 1946) defined the more general formula applicable to equidistance data set as:

$$f(x) = \sum_{k=-\infty}^{\infty} y_k \, \Phi(x - k) \tag{2-22}$$

where Φ was an even basic function. The backbone of any interpolation approach is made of Eq. (2-22) and it was selected as the guide line for choosing an appropriate method. The lip's movement data samples are formulated by a modified version of the Lagrange interpolation. In the following sections, the ideal signal construction, the Lagrange interpolation and Barycentric Lagrange interpolation are introduced.

2.12.1. IDEAL SIGNAL CONSTRUCTION

The other method of signal construction from a sample set, as mentioned, is achieved using the sinc function as the basic function. This method is also referred to as construction of a band-limited signal by ideal low-pass filter. The sinc function was given by (Oppenheim & Schafer, 1999):

$$sinc(\omega_0 t) = \frac{sin(\pi\omega_0 t)}{\pi\omega_0 t} \tag{2-23}$$

where angular frequency of $\omega_0 = 2\pi f_0$ is the fundamental frequency. Eq.(2-23) can be thought as a sinusoidal function $sin(\pi\omega_0 t)$ that has been multiplied by a decaying function $\frac{1}{\pi\omega_0 t}$. The graphical representation of Eq. (2-23) in Figure 2-14(a) visually approves the properties of the sinc function where $sinc(0) = 1$ and follows l'Hopital's rule as (Oppenheim & Schafer, 1999):

$$sinc(\omega_0 t) = 0 \tag{2-24}$$

where $t \in \mathbb{Z}$.

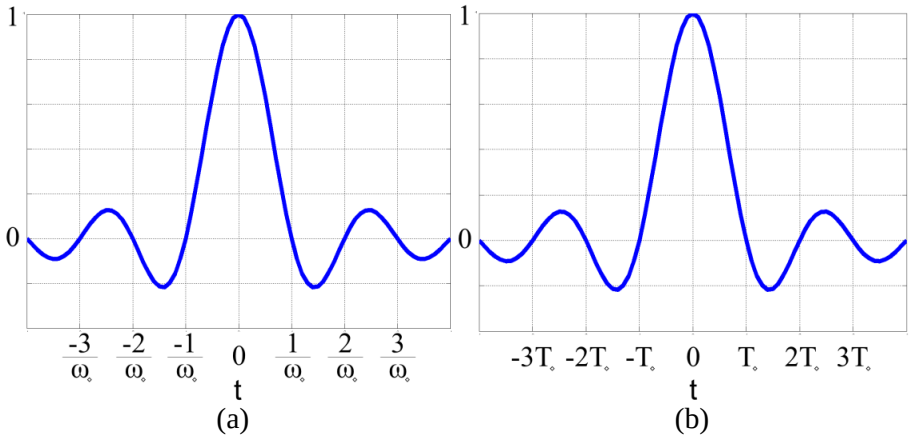


Figure 2-14: The sinc function with (a) normalised (b) and non-normalised domains

In Eq. (2-20) the transfer function for the system in Figure 2-15 is noted by $h[t - nT]$. This expression indicates that each sample is defined in continuous-time domain where distanced from each other by nT.

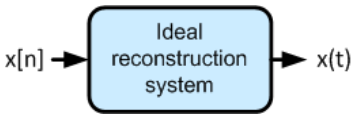


Figure 2-15: The ideal construction of a continuous-time signal $x(t)$ from its discrete-time samples $x[n]$

The freedom of choosing the interval is the main bridge to signal construction and discrete-to-continuous signal conversion. The constructed signal $P_{sinc}(t)$ from a sample set by convolving each sample with the sinc function is (Oppenheim & Schafer, 1999):

$$P_{sinc}(t) = sinc(t) * \sum_{i=0}^{N-1} x(iT_0)\delta(t - iT_0) \tag{2-25}$$

$$= \sum_{i=0}^{N-1} x[i]sinc(t - iT_0) \tag{2-26}$$

where "*" represents the convolution operation.

2.12.2. THE LAGRANGE INTERPOLATION

The Lagrange (Lagrange-Newton) interpolation is a method of polynomial formulation, which constructs a continuous-time polynomial $L(x)$ from N number of samples taken from a function $y = f(x_i)$ and $i \in \mathbb{N}$ over an equidistance interval $x_i \in [0: N-1]$.

The corresponding amplitudes of samples are defined as $a_i = f(x_i), i = \{0, 2, 3, \dots, N-1\}$. The variable x is defined via changing the polynomial's sampling frequency β.

The Lagrange interpolation expressed as (Lagrange, 1877):

$$L(x) = \sum_{i=0}^{N-1} f(x_i)\, l_i(x) \tag{2-27}$$

where $l_i(x)$ is the basic functions corresponding to node $\{x_i\}$ defined as (Lagrange, 1877):

$$l_i(x) = \frac{\prod_{\substack{k=0 \\ k \neq i}}^{N-1} (x - x_k)}{\prod_{\substack{k=0 \\ k \neq i}}^{N-1} (x_i - x_k)} \tag{2-28}$$

where the basic function has the property of (Lagrange, 1877):

$$l_i(x_k) = \begin{cases} 1 & i = k \\ 0 & otherwise \end{cases} \tag{2-29}$$

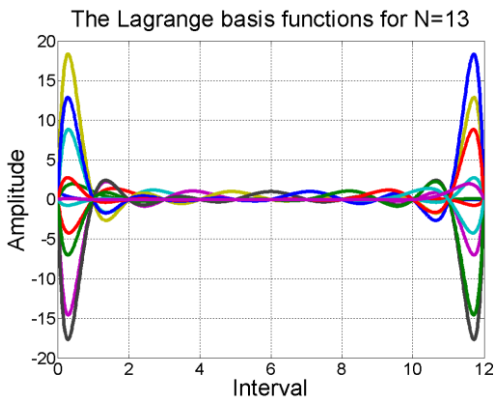


Figure 2-16: The Lagrange basic functions

In Figure 2-16, the basic functions $l_i(x)$ for $N = 13$ and $\beta = 1000$ are shown. The following facts can be observed:

a) Unlike Bezier or B-spline basis functions, the distribution of amplitudes are not even and symmetrical like the.
b) The basic functions amplitudes are affected by the first and the last spans more than the middle of interval.
c) By increasing the number of basic functions, the amplitudes dramatically increase.
d) Like Bezier curve, as the number of data points increases the degree of the Lagrange polynomial could be increased.

Therefore, it is possible to predict the imperfection of polynomial construction.

2.12.3. THE RUNGE'S PHENOMENON

Interpolating a function, with a higher degree polynomial than the function, results in amplitude oscillation at the boundary intervals. This effect is described as the Runge effect (Runge, 1901) and appears in form of oscillation at the ending boundaries. These high amplitude oscillation effects at the end portions of the curve and make the interpolation method impractical when dealing with almost $N > 15$. This effect is shown in Figure 2-17 where a set of 40 random integers with amplitude boundary r, $r \in [1, 10]$ has been fitted with an interpolating polynomial consisting of 1000 samples ($\beta = 1000$). The amplitudes on the first and the last spans of interpolating polynomial are very high ($\approx 3.5 \times 10^8$) and completely dominate middle polynomial's amplitudes. Zooming into its range is the only way to observe the details of interpolating polynomial.

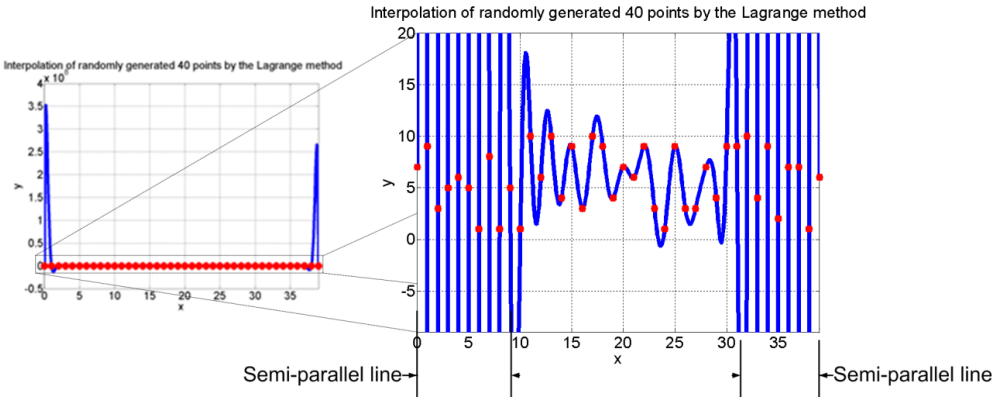


Figure 2-17: An interpolated polynomial by Lagrange method and the details of amplitude

The other drawback of the Lagrange interpolation is that the degree of the interpolated polynomial increases with the number of points used to define it unlike the original function whose degree has not changed.

2.12.4. THE BARYCENTRIC LAGRANGE INTERPOLATION

The Runge's effect can be eliminated by repositioning sample's nodes $\{x_i\}$ and modification of the interval modification given by equation (2-30) (Berrut & Trefethen, 2004) and (Salzer, 1972):

$$L_B(x) = \frac{\sum_{i=0}^{N-1} f_i \frac{w_i}{(x - x_i)}}{\sum_{i=0}^{N-1} \frac{w_i}{(x - x_i)}} \tag{2-30}$$

where the weight function w_i are given by (Salzer, 1972):

$$w_i = (-1)^i \delta_i \qquad \delta_i = \begin{cases} 0.5 & i = 0, i = N-1 \\ 1 & otherwise \end{cases} \tag{2-31}$$

The intervals are defined by Chebyshev points of the second kind. (Berrut & Trefethen, 2004):

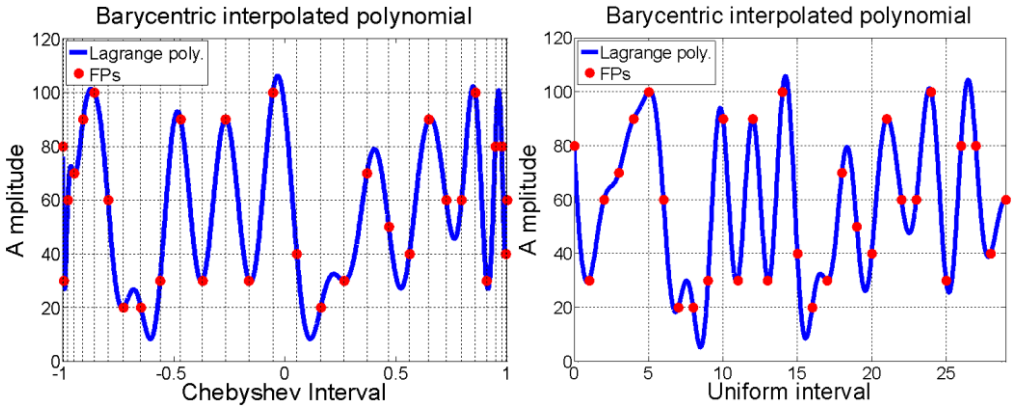
$$x_i{}^{ch} = cos\frac{i\pi}{(N-1)} \qquad i = \{0,1,2,\dots,(N-1)\} \tag{2-32}$$

The Chebyshev points could be considered as the projection of uniformly spanned interval onto $[-1, 1]$. The desirable representation of curve over a uniformly spanned

interval $[a: b]$ is achieved by a transformation that is multiplied to the weight function w_i. It leads (Berrut & Trefethen, 2004):

$$w_j = w_i(\frac{2}{b-a})^{N-1} \tag{2-33}$$

The new weighting function w_j allows the transformation of the polynomial where is defined over interval $[-1: 1]$ to a new interval $[a, b]$. In Figure 2-18 (a), the Barycentric interpolation is applied to 30 samples in the Chebyshev interval and converted to the uniform interval by Eq. (2-33) in Figure 2-18 (b).



(a) (b)

Figure 2-18: The Barycentric Lagrange interpolation (a) over Chebyshev interval, and (b) over uniformly spanned interval

Comparing the Barycentric Lagrange interpolation with Lagrange interpolation shows the following advantages:

a) All samples are included in the interpolating polynomial
b) The Chebyshev points of second kinds discard the high amplitudes oscillation.
c) The intervals can be transformed to any interval.

Therefore, the Barycentric Lagrange interpolation is the best candidate to formulate the visual speech data.

2.13. PULSE WIDTH MODULATION

In this section, the Pulse width Modulation (PWM) is used as the main approach for visual speech barcode generation. The PWM is a well-known method in power

electronics. In this work, amplitudes of a triangle pulse are compared to amplitudes of the BL polynomial. Whenever the amplitude of BL polynomial gets higher than the triangular pulse, the value zero is stored. On the other hand, the value 1 will be allocated to as a bar if the amplitude of BL polynomial gets lower than the triangular pulse.

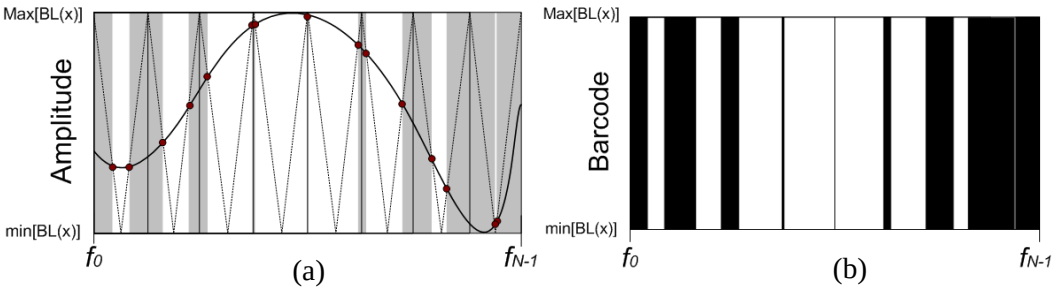


Figure 2-19: The PWM waveform of (a) a signal with triangular pulse train and (b) the resulted barcode

As it shown in Figure 2-19 (a), a typical BL polynomial $BL(x)$ compared with a triangle pulse train over an interval $x \in [f_0 : f_{N-1}]$. In Figure 2-19 (b), result of amplitude detection, arranges the concept of barcode. More details about the procedure will be represented in Section 7.3.

2.14. SOURCE CODING

The lips are source of all possible visual speech signals that could be generated. It is possible to assume all the generated visual speech signals are combinations of limited symbols. In this section, concepts of signal quantization and coding are introduced. The reason of quantizing signal is to reduce the variation of signal. It provides statistical information for the next level that is entropy coding. In this work, the Huffman approach has been chosen for coding the quantized signals.

2.14.1. SIGNAL QUANTIZATION

Quantization of a signal could be thought as rearranging and shaping the samples amplitudes to certain quantities. Quantization is a process, which divides the range of a signal to different regions, and allocates equal values to all samples within the regions. The regions, which are defined over input signal, are called decision levels while the values of the quantized signal are called construction levels (Mandal & Asif, 2007). Since it maps all the samples in limited regions to wider boundaries than main

amplitudes, some information would be lost. In this work, a decision level is referred by a partition. The decision levels are noted by $P_i, i = 0,1,2,\dots,l-1$ where the quantizers's number of levels is equal to l. The difference between decisions $\Delta = P_{l+1} - P_l$ is called quantization step size. Each construction level $L_q, q = 0,1,2,\dots,l-1$ is located between the decision levels ($P_l < L_q < P_{l+1}$).

Based on the arrangement of quantization step size and construction level, there are uniform and non-uniform quantizers as they shown in Figure 2-20 (a) and (b), respectively. The equivalent representation of uniform and non-uniform quantizers could also be obtained by allocating the x axis to the input signal and simultaneously representing the decision and construction levels on y axes as they are illustrated in Figure 2-20 (c) and (d). The decision levels in uniform quantizer are obtained by $L_q = \frac{P_l + P_{l-1}}{2}$. On the other hand, the non-uniform quantizer has variable step sizes.

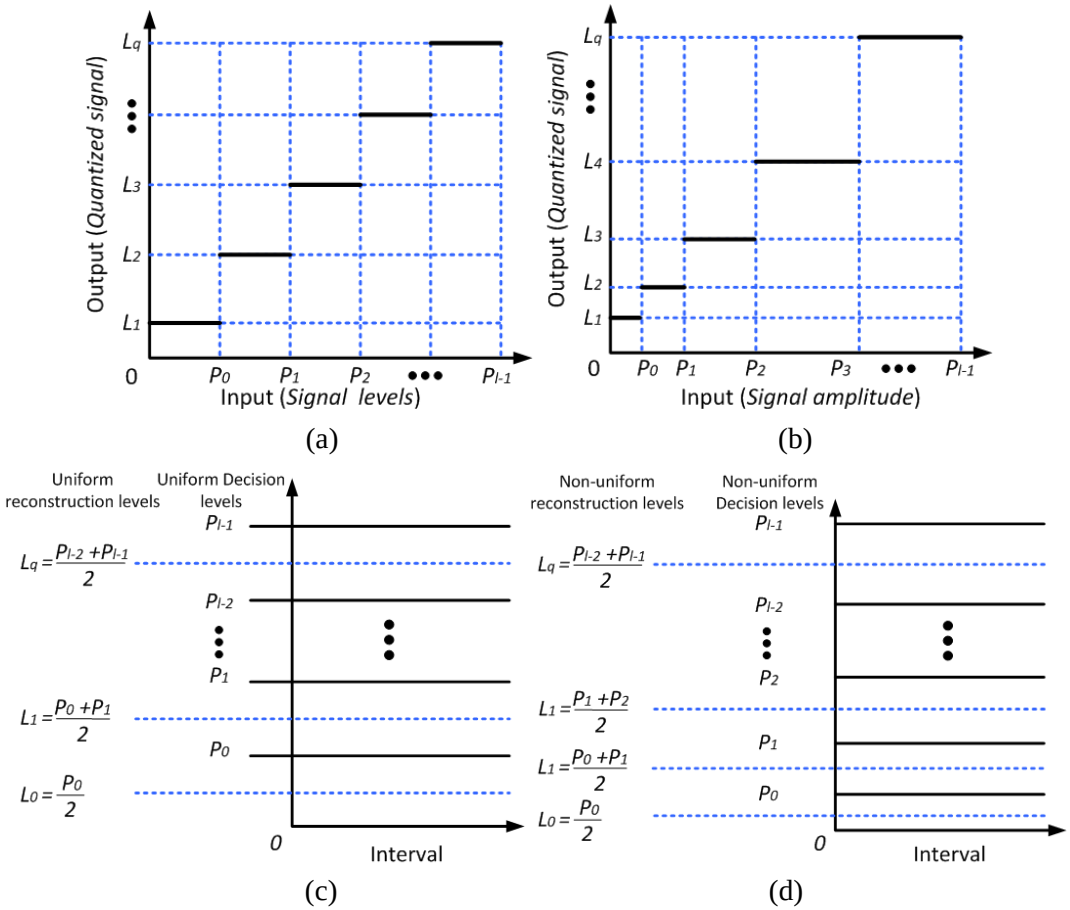
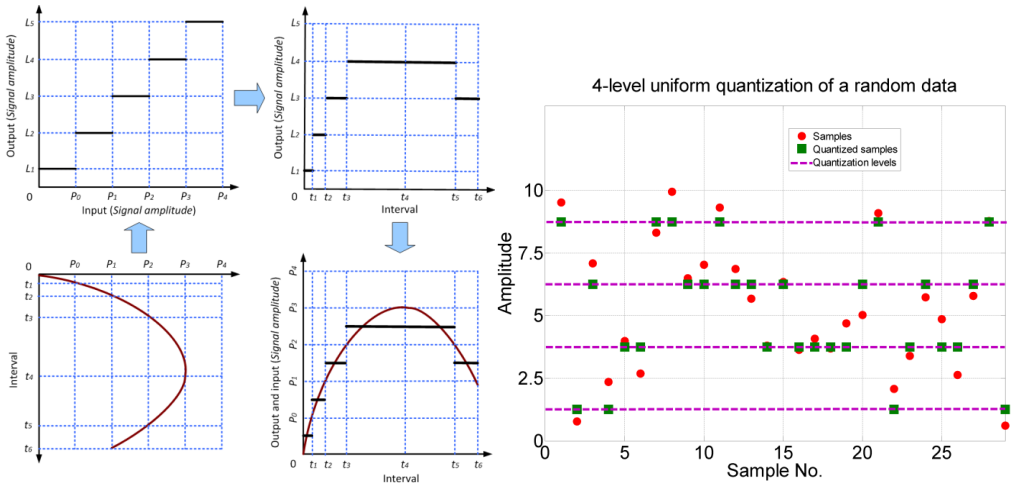


Figure 2-20: The (a) input-output relation of L=5 level uniform (b) and non-uniform quantizer as well as the equivalent representations in (c) and (d), respectively

The operation of quantization on a signal's amplitudes $x[n], n = 0, 1, 2, \dots, N-1$ for non-uniform quantization is denoted by $x_q[n]$ and defines:

$$x_q[n] = \begin{cases} L_0 & if \quad 0 \le x[n] < P_0 \\ L_1 & if \quad P_0 \le x[n] < P_1 \\ L_2 & if \quad P_1 \le x[n] < P_2 \\ \cdot & \cdot \quad\quad \cdot \\ \cdot & \cdot \quad\quad \cdot \\ \cdot & \cdot \quad\quad \cdot \\ L_q & if \quad P_{l-1} \le x[n] < P_l \end{cases} \qquad n = \{0,1,2, \dots, (N-1)\} \qquad (2\text{-}34)$$

In Figure 2-21 (a), the quantization process of a signal is shown. This process inputs a signal, which has been bounded by a set of partitions, and allocates a particular level to the corresponding levels. Each boundary is defined a non-uniform spans in signal domain. The quantized signal is defined over such time span and preserves an approximation of signal's amplitude.



(a) (b)

Figure 2-21: The process of (a) signal quantization applied to the (b) uniform quantization and 4-level quantization of a set of random data

For instance, in Figure 2-21 (b) a randomly generated signal from uniform distribution with the length of $N = 30$ and bounded between zero and one is uniformly quantized.

- Quantization error

The process of quantization is not reversible. It means the information about the input signal cannot be preserved by quantized signal. The amount of relation between input

signal $x[n]$ and quantized signal $x_q[n]$ can be expressed by subtracting them. The quantization error is obtained as the result of subtraction:

$$e_q[n] = x_q[n] - x[n] \tag{2-35}$$

The quantization S/N ratio is obtained by $(S/N)_q$ (Oppenheim & Schafer, 1999):

$$(S/N)_q = 10log(\frac{\sum_{n=0}^{N-1}|x[n]|^2}{\sum_{n=0}^{N-1}|e[n]|^2}) \tag{2-36}$$

2.14.2. HUFFMAN CODING

The Huffman coding is an entropy-coding scheme for digital signal compression. This coding scheme allocates lower number of bits to the more frequently occurring values and higher number of bits to the less frequently occurring values in a digital signal. This characteristic implies to the usage of statistical information about the digital signal values. The Huffman dictionary consists of all possible amplitudes emit from signal source which is called coding symbols.

The Huffman coding has the following characteristics:

a) The number of symbols in the Huffman coding scheme is N_m (Huffman, 1952). The probability of each symbol $P_s(i)$ is (Huffman, 1952):

$$\sum_{i=1}^{N_m} P_s(i) = 1 \tag{2-37}$$

b) The number of messages N_m can be replaced by the value of each level in the quantized signal L_q.
c) Each symbol has a binary value of code noted as 'code word'.
d) The collection of code words is called code dictionary
e) The Huffman coding scheme uses statistical properties of entire amplitudes for choosing the symbols.
f) The length of each message is $L(i)$ and the average length is $L_{av} = \sum_{i=1}^{N_m} L(i)\, P_s(i)$.
g) The information carried by each message is denoted by $I(i) = -log_2 P_s(i)$.
h) The message redundancy $H(i)$ is the expected value of message information. Therefore, it can be calculated as (Huffman, 1952):

$$H(i) = -\sum_{i=1}^{N_m} P_s(i)\, log_2 P_s(i) \tag{2-38}$$

i) The longer code increases the coding efficiency but it comes with the price of computational cost and memory capacity.

In this work, Huffman codding has been chosen for representing digital visual speech signals.

2.15. SUMMARY

In this chapter, the fundamental definitions and terminologies have been followed by introducing the field of study in visual interpretation of speech. Afterward, the potential mathematical models for representing the lip dynamism during articulation have been introduced. The BLI method has been chosen as the best mathematical candidate for formulating the visual speech information. For further simplifications of visual speech signals, the PWM and source coding have been suggested.

Chapter 3

DESIGN OF THE CORPUS

In this work, during study of the phonemic relations with words, it has been found that deriving signatures for visual speech signals will be more comprehensive if it has been produce from a specific family of words. Furthermore, allocating the signature representation for the visual speech signal when it lacks of systematic structural relations is hardly possible. Therefore, the structural relations need to be detected in the words linguistic domain.

The three phonemes sequences, which appear in the transcribed words, are determined by database hierarchical analysis. The sequences of phonemic symbols are the subject of such analysis. By concluding the most occurring phonemic alphabets, the most occurring set of words are extracted. The phonemic structure makes the words families.

In this work, studying the behaviour of words phoneme sequence is called phoneme-based analysis of words. In this chapter, the method of designing the transcribed corpuses is describes. The process was depicted earlier in Figure 1-2. As mentioned in the literature review, the articulating lip carries similar (bimodal) information to audible speech. Processing the audio information is substituted by processing the transcription of words. In other words, the visual information is interacting with the phonemic structure of words.

The visual speech signals and their processed versions form a group that considered as visual words. Extracting the lip's motion during the uttering of the chosen words is the first step for obtaining the visual speech signal.

3.1. PHONEMIC STRUCTURE OF WORDS

In practice, words have specific sequences of phonemes that uniquely define them. The relation between phonemic sequences and words is the index of grouping words with same phonemic foundation. Clearly, categorising the most reliable combination of phoneme sequence must start from the beginning of words. It means there is no need to study the random combination of phonemes. Obviously, many words sharing the same phoneme sequence in their beginning. This concept looks like growing a peach tree from a seed or lighting from a particular part of cloud. The peach tree's body is the origin of branches and also they consist of other branches that hold the fruits. One of the significant advantages of phoneme-based analysis of words is ability of controlling the mathematical expressions in the visual domain. In this work, searching for common phonemes in words starts for first three sequences of phonemes (triphones).

There is a specific phonemic sequence where each phone succeeds by next phone forming the final utterance. In a population of words, these specific patterns are shared in word's different sections such as in the beginning, in the middle, and in the end. It means there are some words that start with the same combination of phonemes or some share the same phonemic sequences in the middle and some in the end. These fixed phonemes structures affect the articulation of each individual phoneme and includes the Co-articulation effect. The similar effects in visual domain (combination of visual phonemes) are happened on the corresponding sequences of the visual features of lip's movement in frame domain. Therefore, finding the meaningful sets of phonemes is the first step in studying the Co-articulation effect. The phonemic structure of words defines these sets. Finding a section of word in which the phoneme sequence appears more recognizable is the second task.

Extracting the words phonemic rules starts with adopting a collection of transcribed words containing a reasonable numbers of words. The other important issue is the source of words collection. More specifically, selecting the words by their phonemic structure of a reliable vocabulary database enriches the findings while associating systematic categorization of words.

For example, consider the words ‘simple’, ‘symptom’ and ‘simmer’ with phonemic representation as /s/+/ih/+/m/+/p/+/el/, /s/+/ih/+/m/+/p/+/t/+/ax/+/m/ and /s/+/ih/+/m/+/ax/, respectively. The three specific phonemes sequence /s/+/ih/+/m/ is in common between these words while in the fourth position of phonemic representation, ‘simple’ and ‘symptom’ share the phoneme /p/.

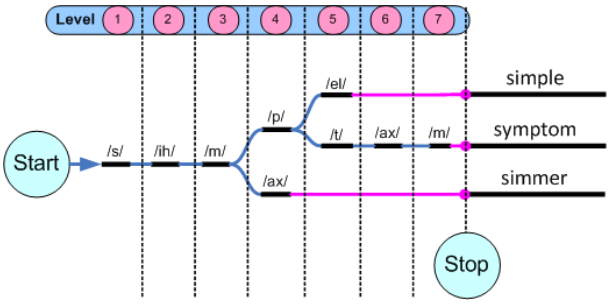


Figure 3-1: Designing the corpus by tree representation

This is the basic method of categorizing the words in a tree diagram representation as shown in Figure 3-1. This concept proves the relation between the mathematically expressed visual speech signals called visual word and the meaningful combinations of sounds. Consequently, the mathematical expressions, which would be derived for the visual words, are categorized into a group. Therefore, the words are grouped mathematically. In mathematical form, the difference between each member of group will be revealed after the third phoneme. Referring to Figure 3-1, the lip’s movement (this time in frame domain) would be same till the third level while after the fourth level if the visual features corresponded to phoneme /p/ is received, the word ‘Simmer’ left out of the results. Similarly, if the visual features in fifth level are detected, the word ‘Simple’ would be detected and there is no need to look forward for analysing the other feature points.

Therefore, the visual speech signatures demonstrate a systematic relation between a member of word’s family and the other words family members. This method provides an interesting ability to find how the mathematical expressions could be changed to form other words in the group although it is not the subject of studying in this work. This method could be used in lip-reading and also animation systems. Toward establishing the signature expression for structural words, a set of words are chosen by the criteria of maximum phonemic sequence occurring among all the words transcribed text. This method can be applied to any transcribe dictionary however, due

to time restrictions, groups of words are chosen in English language for synthesizing the visual word database.

3.2. CONTEXT DATA PREPARATION

Phoneme consists of consonant and vowel denoted by C and V, respectively. Categorizing a collection of words according to phonemes could reveal the phones popularity in human usage contest. Therefore, finding a limited set of words appearing with maximum probability in a collection of a vocabulary database is the aim of phoneme structural analysis. Such manner of categorization also reveals the relation between phonemes and its visual correspondence for word generation and perception. The signature concept, which is supported by such properties, will also follow a structural behaviour. Furthermore, the knowledge about studying such relation includes the Co-articulation model of uttering words that also plays a crucial role in the visual speech domain. For the audio-visual databases, which could be used using in audio-visual speech recognition systems, knowing how many words start with consonant or vowels and are followed by other consonants and vowels will narrow down the issue of searching. It means the employed searching algorithm does not need to examine the entire database to interpret the extracted visual speech signals (see Figure 3-1).

It should be mentioned that the vast amount of data needs to be organized in the phonemic structure, which increases the computational cost but the result would be a breakthrough for new generation of lip-reading and animation systems. Furthermore, the grammatical rules, e.g. past participle versions of the words need to be considered. Based on these facts, only a group of words are designed according to population properties of phonemes. The population properties of phonemes as they appear sequentially will be described in the following sections.

In conclusion, the suggested design for transcribed corpus is formed by answering the following questions:

1 . Which corpus can be a good candidate?
2 . How many phonemes exist in a population of the corpus?

3 . Which phonemes appear with the maximum population for three sequences of phonemes (triphone)?
4 . What are the words (final corpus) who share such triphone?

The answers to these questions will be provided in the following sections.

3.3. PHONEMIC-BASE ANALYSIS OF WORDS

The analysis of phonemes sequence has been conducted on the TIMIT audio-visual database corpus. The database of text corpus along with its phonemic structure is used for extracting a family of words for the aims of this research. Toward acquisition of words structures, the hierarchical analysis based on the population of phonemes is employed for extracting information about the appearance frequency of phoneme in the text corpus. According to the position of phoneme starting in the beginning of entire transcribed text, in three stages, phoneme sequences with maximum occurrence are determined. This specific configuration of phonemes includes a limited number of words. This method of word extraction in the higher level can be extended to the complete set of words but it is out of the scope of this work. Toward the goals of this work and the mentioned advantages, in next sections the words phonemic structures are analysed from the transcribed text in an audio-visual database corpus and will be used for visual feature sample sets extraction.

3.3.1. SELECTING A PHONEME-VISEME TABLE

In the field of speech processing, it is possible to select the appropriate viseme table that has been selected as the visual domain representation reference. Selecting appropriate phoneme-viseme mapping table for analysis of phonemic structure would be beneficial in deriving the practical visual speech signal expressions that are adaptable to audio-visual speech recognition systems. In other words, the method of deriving visual speech signal signature can be applied to the speech recognizer system as its collaborating visual domain. According to (Potamianos et al., 2004) a phoneme set used in Hidden Markov Model Toolkit (HTK) is selected as the phoneme-viseme mapping table. In this mapping scheme, which is represented in Table 3-1, the many-to-one strategy is applied based on the manner of articulation.

Table 3-1: The phoneme-viseme table (Potamianos, et al., 2004), phonemes are grouped into viseme classes

	Consonants	Vowels
1	/p, b, m/	/ih, iy, ix, ax/
2	/f, v/	/ae, eh, ey, ay/
3	/l, el, r, y/	/uw, uh, ow/
4	/s, z/	/ao, ah, aa, er, oy, aw, hh/
5	/sh, zh, ch, jh/	
6	/t, d, n, en/	
7	/th, dh/	
8	/ng, k, g, w/	

The consonants consist of eight subgroups of phonemes that have visual similarities during articulation while, vowels are categorized into four subgroups. Although the English alphabet consists of 26 letters, there are 43 phonemes. This fact shows the difference between possible sound in words and the letters. Now it is the time to find the transcribed words text that is marked according to this phonemic table.

3.3.2. SELECTING THE CORPUS FOR DESIGNING A LIMITED SET OF WORDS

In order to design a phoneme-based analysis of words, a transcribed vocabulary database should be selected and analysed. The analysis has been conducted on the TIMIT's vocabulary database. In this study, the TIMIT's transcribed dictionary with corresponding phonemic notations was used (Garofolo, et al., 1993). This database was mainly designed for speech recognition system but it was also employed in lip-reading system (Potamianos, et al., 2004). The block diagram in Figure 3-2 shows more details about the structure of the chosen database as well as an adaptation phase based on the chosen phoneme-viseme mapping table.

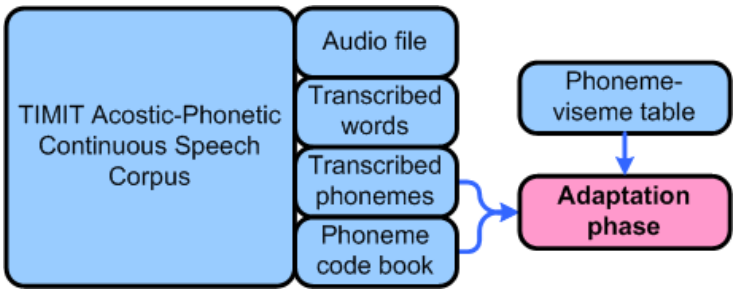


Figure 3-2: Relating the TIMIT components to the phoneme-viseme table

For analysis of the words' phonemic structure, three sets of information as TIMIT's transcribed words, transcribed phonemes, and phoneme codebook are used although the adaptation process is only applied to transcribed phonemes, and phoneme codebook. Using a phoneme-viseme table and analysing the phoneme sequential structures in the dictionary can be considered as bridging between speech recognition and lip-reading systems. The adaptation phase will be described in the next section.

3.3.3. ADAPTATION OF THE DATABASE TO THE PHONEME-VISEME TABLE

There are phonemic contexts of words in the TIMIT's dictionary that have been marked with $\{1, 2\}$ representing the position of the stress. For example, the first stress for the word is marked by adding 1 to phoneme /ey/ that leads to /ey1 b el/. These numbers are eliminated in the modified TIMIT's dictionary for simplicity. In addition, only the phonemes that are represented in Table 3-1 are the subject of words phonemic structure analysis.

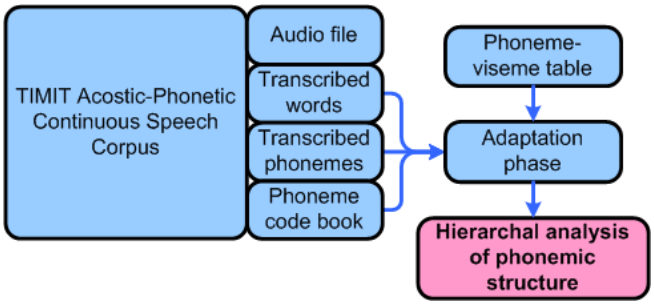


Figure 3-3: The outcome of adaptation phase

The main adaptation occurred by comparing the TIMIT's phonemic symbols in Table 2-3 and Table 3-1. It means the words that only are defined by phoneme groups in Table 3-1are the subject of processing. One of the effects of using the phoneme-viseme table is the reduction of number of words that are described with the TIMIT's phone code. By adapting the phoneme-viseme table and the transcribed database, the hierarchal analysis of phonemic structure is determined and described in the following sections.

3.3.4. HIERARCHICAL PHONEME-BASE ANALYSIS

In this part, the hierarchical analysis is used for studying the combination of phonemic components in word. In the hierarchical analysis of words phonemic structure, the

goal is grouping all possible sequential appearance of consonants and vowels. According to (Seels & Glasgow, 1990) "A hierarchy was an organization of elements that, according to prerequisite relationships, describes the path of experiences a learner must take to achieve any single behaviour that appears higher in the hierarchy". Applying this statement to this work, the task is to find the groups of words by studying the sequential appearance of phonemes. This process is shown in Figure 3-4.

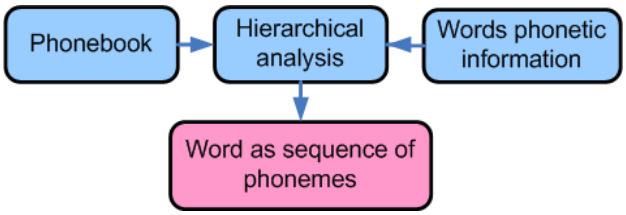


Figure 3-4: Searching for word's phonemic rules

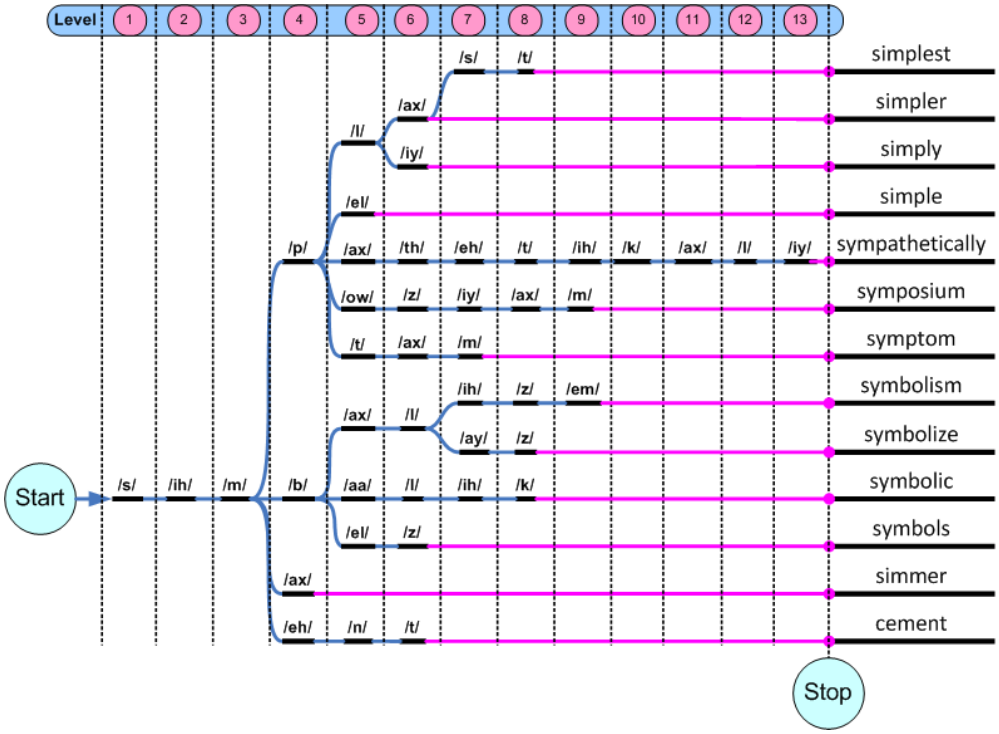
The outcomes appear as the probability of phoneme existence is determined. A decision tree visualized the hierarchy and represents the frequency of words that contained the consonants and the vowels in the first phoneme positions of the words. In the second and third level, other branches of tree are determined according to appearing phoneme in the words.

3.4. WORDS PHONEMIC PATTERN ANALYSIS BY DECISION TREE

The decision tree provides the information for systematic analysis of the organized data. By employing this approach to the phonemic representation of the words, the hierarchical probability information of phonemes sequences as they proceed in words are extracted. A decision tree, which is defined in sequentially arranged levels, consists of node. The nodes relations are visualized by connecting line from their current level to other nodes for the successive level. The nodes represent a particular state, group, event, or position in each level that exist in the data set. The connecting lines represent conditional rules that define the transition amongst the nodes between two consecutive levels. These rules could be stated by probabilistic or occurrence information about the next state of the nodes. In this work, the nodes of the decision tree are allocated to the C and V groups. At the output of each node there are two possible outcomes. The connecting lines are showing the probability of transition between Cs and Vs on sequences of levels $l_n, n \in \mathbb{N}$ with n that stands for the number

of levels. In this way, an n-level tree defines 2^n possible paths that represent C and V sequences (see Figure 3-13).

Another type of a tree representation of data is applied to the final result of words grouping process which is depicted in Figure 3-5. In this version of decision tree, each node is allocated to a possible phoneme, which has been detected in each level.



(a)

Level	1	2	3	4	5	6	7	8	9	10	11	12	13
Defined words	0	0	0	1	1	4	1	3	2	0	0	0	1

(b)

Figure 3-5: The tree representation of (a) phonemes sequences categorizing all thirteen words as branches of one root and (b) the number of detected words in each level

The words have same phonemic sequence for three levels, which occurred with maximum population in each level. This trend continues for thirteen levels and the final word is defined Figure 3-5 (a). The numbers of detected word are also represented in Figure 3-5 (b). In the level six, the maximum four numbers of words are fully categorized in this group of words.

3.4.1. METHOD OF CONTEXT ANALYSIS

Searching the phonemes in the TIMIT's transcribed dictionary is performed using the Microsoft Word processor. Using the 'find and replace' option makes finding the phonemic structure of words by following the TIMIT's phone codes possible. Each word's phonemics representation is located between two slashes. For instance, the word 'celebrate' is represented phonemically by /s-eh-l-ix-b-r-ey-t/. Note that the dashes are representing spaces. Therefore, if someone needs to find all the words starting with /s-eh-l-…/ phonemic structure, he will need to search for '/s-eh-l-'. The results of searching are the number of words that are following the particular sequence of phonemes provided in the dialogue box. From the previous example, the total numbers of 12 words sharing such root are detected. It is very important to use the spacing character properly in the searching process. Analysis of phonemes sequential structures in TIMIT's dictionary has been conducted by using this method.

3.4.2. ANALYSIS OF THE CONSONANT AND VOWEL

Searching for C or V in the dictionary, which is represented in Figure 3-6 is conducted on two phases. In the first phase, examination is conducted on the whole dictionary, regardless to the positions of phonemes. This provides the basic information about the phonemic population in the dictionary in three main positions. From this point of view, the highest Cs or Vs used are detected in the beginning, middle and ending of words.

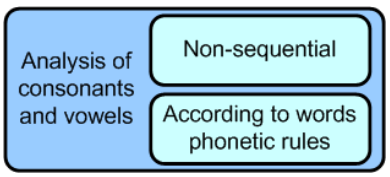


Figure 3-6: Possibilities in searching phonemes the vocabulary database

In the second phase, the sequential appearance of Cs and Vs are considered. The sequential analysis provides the fundamental details about arrangement of phonemes where each word follows.

3.4.3. NON-SEQUENTIAL

The appearance of phonemes in the TIMIT's vocabulary context will give basic understanding about the words phonemic structures. Studying the position of consonants or vowels in the beginning, the middle, and the end of all words, reveals more comprehensive understanding about the dictionary that will be formed. This knowledge will be useful for designing possible classification and searching algorithms in lip-reading systems. The outcomes help designing the searching algorithms in practical applications. For example, in a search engine of a lip reading system, if a phoneme does not appear in the beginning of vocabulary context, further examinations will be eliminated. Furthermore, in this work, this process revealed two important characteristics of the dictionary by providing the overall population of phonemes as well as the average length of the words.

Table 3-2: The number of consonant and vowel phonemes appearing in the beginning, middle, and ending of TIMIT's transcribed dictionary

	C	V	Total
In the beginning	4784	1440	6224
In the middle	12960	11224	24184
In the end	5014	1179	6193
Total	22758	13843	
	36601		

In Table 3-3 the number of consonants and vowels as they appeared in three different positions of the words are represented. The number of consonants in all three positions is higher than the number of vowels. In Figure 3-7, the details of phonemic population in the TIMIT dictionary are represented. Furthermore, the total number of phonemes in the beginning and the ending of words must be equal but due to usage of external phoneme-viseme table (Table 3-1), these values are not equal (see Section 3.3.2). It means 31 words are neglected.

In Figure 3-7 more details about the population of phonemes in the three parts of words are represented. The phoneme /s/ is appeared 721 times in the beginnings of the words while the consonant /n/ appeared 1612 times in the middle positions of the words. In Figure 3-7, 915 words finish with the articulating consonant /z/. In Figure 3-7(a), the phonemes /el/, /zh/, /en/, /ng/, /uh/ and /ix/ did not appear in the beginning

of any word. In this early stage, it can be claimed that the final possible corpus consists of 37 main groups.

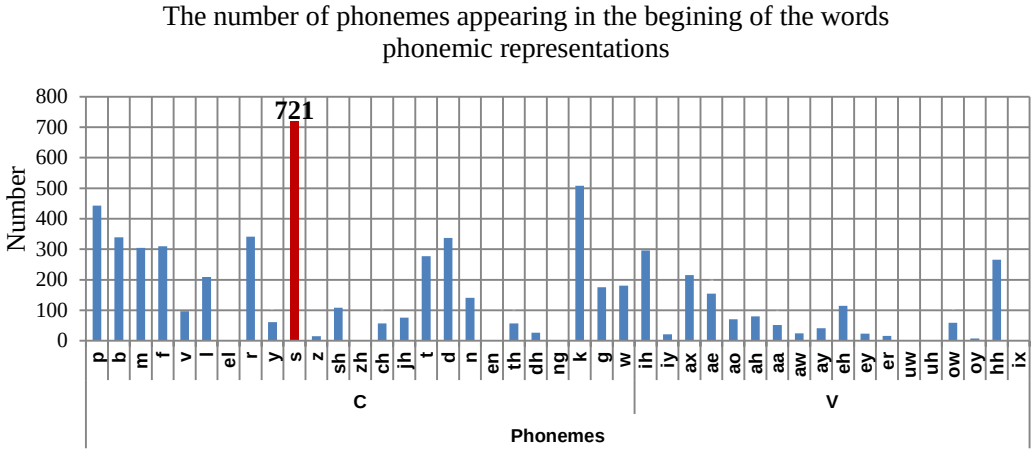


Figure 3-7: The phonemic populations in (a) the beginning, (b) middle, (c) and ending of TIMIT's dictionary

It is also possible to conclude the percentage of consonants and vowel populations according to Table 3-2. Figure 3-8 (a) shows such proportions as the words start in the

first level; approximately 23.2% of them are represented by vowels while 76.8% are represented by consonants. In the middle of the words, 46.41% of phonemes are vowels and 53.59 % are consonants as it represented in Figure 3-8 (b). In Figure 3-8 (c), the 80.26% of the words finish with consonants and 19.74% of them finish with vowels.

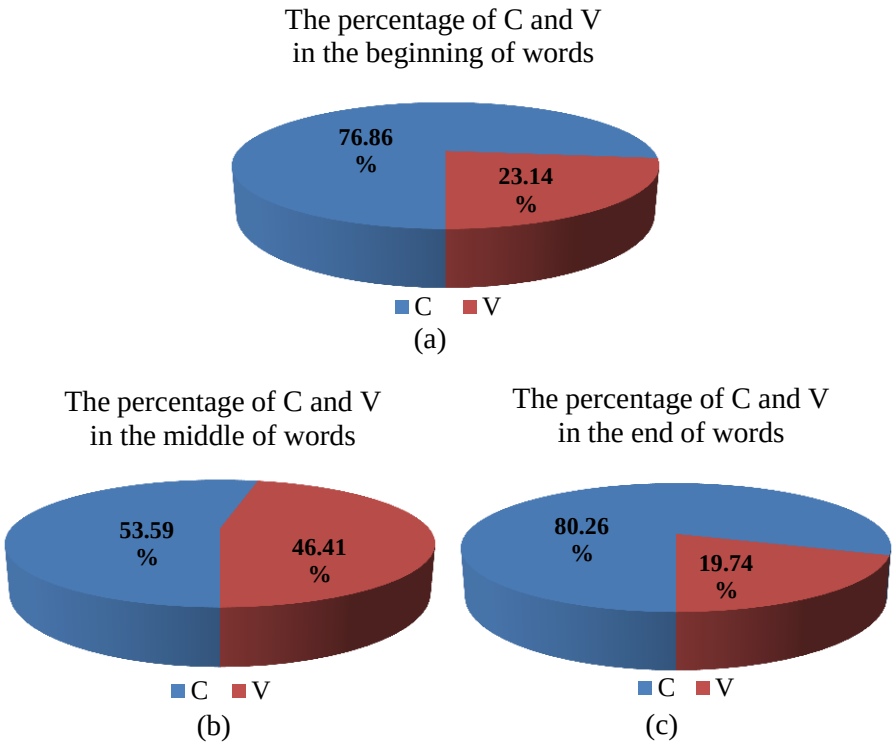


Figure 3-8: The percentage of consonant and vowels in (a) the beginning, in (b) the middle, in (c) the end of the TIMIT's transcribed corpus

According to Figure 3-9 (a), the total percentages of C and V regardless to their position in the transcription are 62.09% and 37.91%, respectively. Conclusively, the total percentage of phonemes appearance in the dictionary is shown in Figure 3-9.

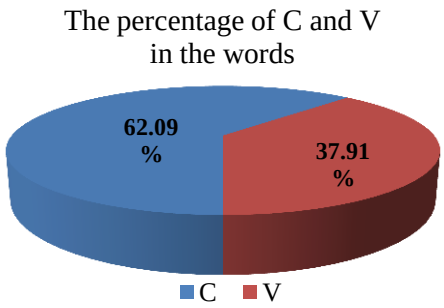


Figure 3-9: The total percentages of C and V regardless to their position in the TIMIT's transcript database

By increasing the number of levels more phonemic patterns of words would be determined while computational cost also increases. Obviously, the number of levels n in a tree does not tend to be infinite and finishes according to the length of the longest word(s). Since obtaining more information about the statistical information of phonemes in the structure of words is tremendously time consuming, in this work only the fundamental of phoneme structures are analysed. In this stage, to obtain an approximation of the number of levels, the average phonemic length of words should be determined. In other words, the goal is to find the average words' lengths when they are uniformly distributed with the same number of phonemes. The average length of the words is simply found by calculating the number of phonemes in the vocabulary set but the gaps between phonemic alphabets should be noticed. The relation of phonemes and gaps can be explained by a simple example. Assume two words are detected consisting $P = 17$ phonemes and there is no information about their actual lengths. By simply dividing the number of phonemes to the number of words, an average length of $P_{ave} = 8.5$ will be obtained. This means it is possible to assume one word consists of eight and the other nine phonemes. In case of two or more words with even numbers of phonemes, for instance, $P = 12$, the average length is $P_{ave} = 6$. This means there are five gaps in the phonemes in average.

Referring to the words in TIMIT's dictionary, there are $W_{TIMIT} = 6224$ transcribed words. All the words are phonemically described.

$$P_{ave} = \frac{P}{W_{TIMIT}} \tag{3-1}$$

The total number of extracted phonemes is $P = 36601$ (see Table 3-2). Substituting values for P and W_{TIMIT} into Eq. (3-1), the average length of a word in the TIMIT's dictionary is $P_{ave} = 5.88$. It can be assumed that if the distribution of lengths is uniform, a tree structure with 6 levels will cover 50% of all words.

3.4.4. WORDS PHONEMIC RULES

After finding the population of phonemes in different positions of TIMIT's vocabulary transcript, the sequential searching process starts by categorizing the words base on their starting phonemes. From this part forward, the combinations of phonemes are categorized and the decision tree would show a map of words until the

third phoneme position. The modified TIMIT's vocabulary context is used for determining the sequence of C and V that constructs phonemic clustering.

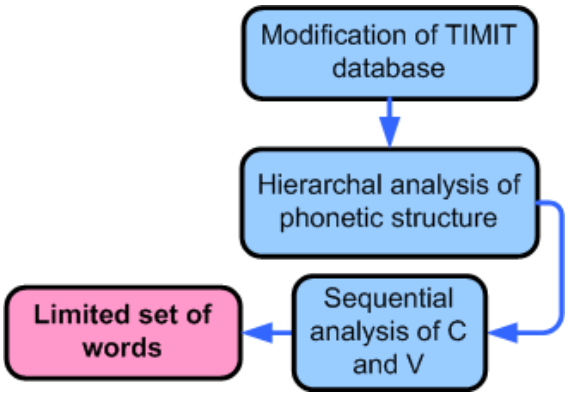


Figure 3-10: Extracting the words by their phonemic representation

The process of obtaining sequential phonemic structure is depicted in Figure 3-10. After modifying the TIMIT's database, hierarchical analysis of phoneme clusters the words based on their phonemic structures. Obtaining the maximum repeated phonemes starting from the beginning of words until the third phoneme leads to finding a group of words that share the maximum number of phonemes that exist in the TIMIT's vocabulary database.

The phonemes in the first positions of words are the same as those which are detected in the non-sequential analysis (see Table 3-2). The number of all possible words, which start with combinations of C and V in the first two phonemic places, are represented in Table 3-3 and the percentages of them are shown in Figure 3-11 (a). For the next level, the number of possible words, which start with C and V in the first three phonemic positions, are represented in Table 3-4 and their percentages are represented in Figure 3-11 (b).

Table 3-3: The number of /CC…/, /CV…/, /VC…/ and /VV…/ phoneme combinations in the TIMIT's dictionary

	C and V combinations in second level			
	/CC…/	/CV…/	/VC…/	/VV…/
Number of covered words	1023	3759	1164	163
Total	6109			
Percentage	21.39	78.61	87.72	12.28

Note that after finding the 2nd and the 3rd levels, some words already have been fully described. For example, the words ‘go, /CV/’, ‘you, /VV/’, ‘of, /VC/’ are found in the 2nd stage and the words ‘oily, /VCC/’, ‘car, /CVC/’ in the 3rd stage. These are not the desirable words since the pattern of maximum phoneme probability does not follow the overall maximum branch. On the other word, they do not last more than three phonemes.

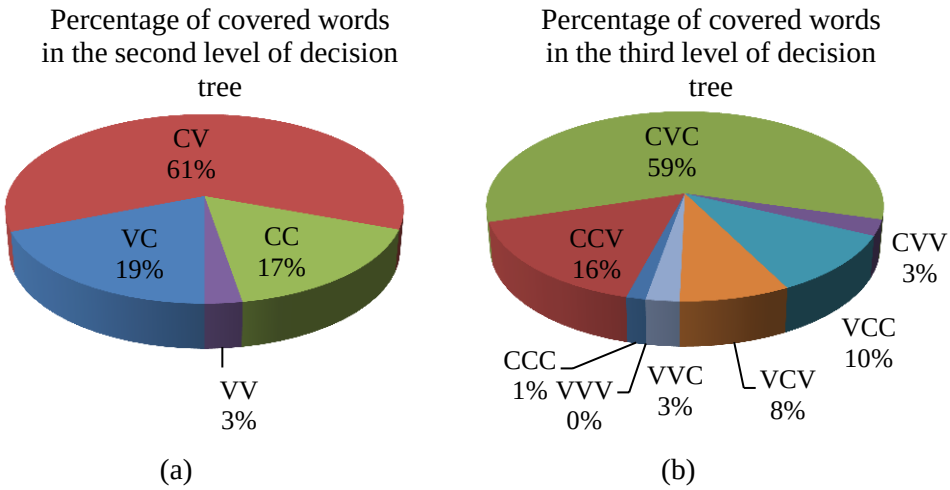
Table 3-4: The number of /CCC…/, /CCV…/, /CVC…/, /CVV…/, /VCC…/, /VCV…/, /VVC…/ and /VVV/ phoneme combinations in the TIMIT’s dictionary

	C and V Combinations							
	/CCC…/	/CCV…/	/CVC…/	/CVV…/	/VCC…/	/VCV…/	/VVC…/	/VVV…/
Number of covered words	82	941	3466	159	610	464	141	0
Total	5863							
Percentage	8	92	95.6	4.3	56.7	43.3	100	0

Table 3-5: The number and percentage of words with maximum appearance according to the phonemic structure in each level of searching

	l_1	l_2	l_3
	/C…/	/CV…/	/CVC…/
Number of words	4783	3759	3466
Percentage of covering the transcribed words	76.86	61.53	59.1

Determination of a path, which has the maximum probability of generating sequential phoneme, reveals the highest numbers of word in the context of C and V combinations (see Figure 3-13). Therefore, in the next stage the /CCC…/, /CCV…/, /CVC…/, /CVV…/, /VCC…/, /VCV…/, /VVC…/, /VVV…/ structures are examined.



(a) (b)

Figure 3-11: The pie charts of the number of covered words according to Table 3-3b(a) and Table 3-4 (b)

In Table 3-5, the number of existing words according to the highlighted path in Figure 2-13 is represented for each level. In this stage, the overall phonemic path can be found. The phoneme sequence /CVC…/ is detected with the most common root for the words.

In Figure 3-12 (a) and Figure 3-12 (b), the details of phoneme population in the second and third positions are shown with more details. In the second level, the phoneme /ih/ has the maximum occurrence of 61 among other vowels. In the third level, the phoneme /m/ had maximum rate in 13 words. It is very important to mention the reason for choosing the vowel /ih/ as the next successor of consonant /s/ although the consonant /t/ occurs 125 times in Figure 3-12 (a). It can be observed that the overall number of vowels appearing in the second level is higher than the consonants. Therefore, the representative of the phoneme in this level is chosen as /ih/, which is the highest among other vowels. The phoneme /m/ with the population of 13 is chosen as the third and last phoneme for grouping words. Note that its repetition also indicates the final number of words.

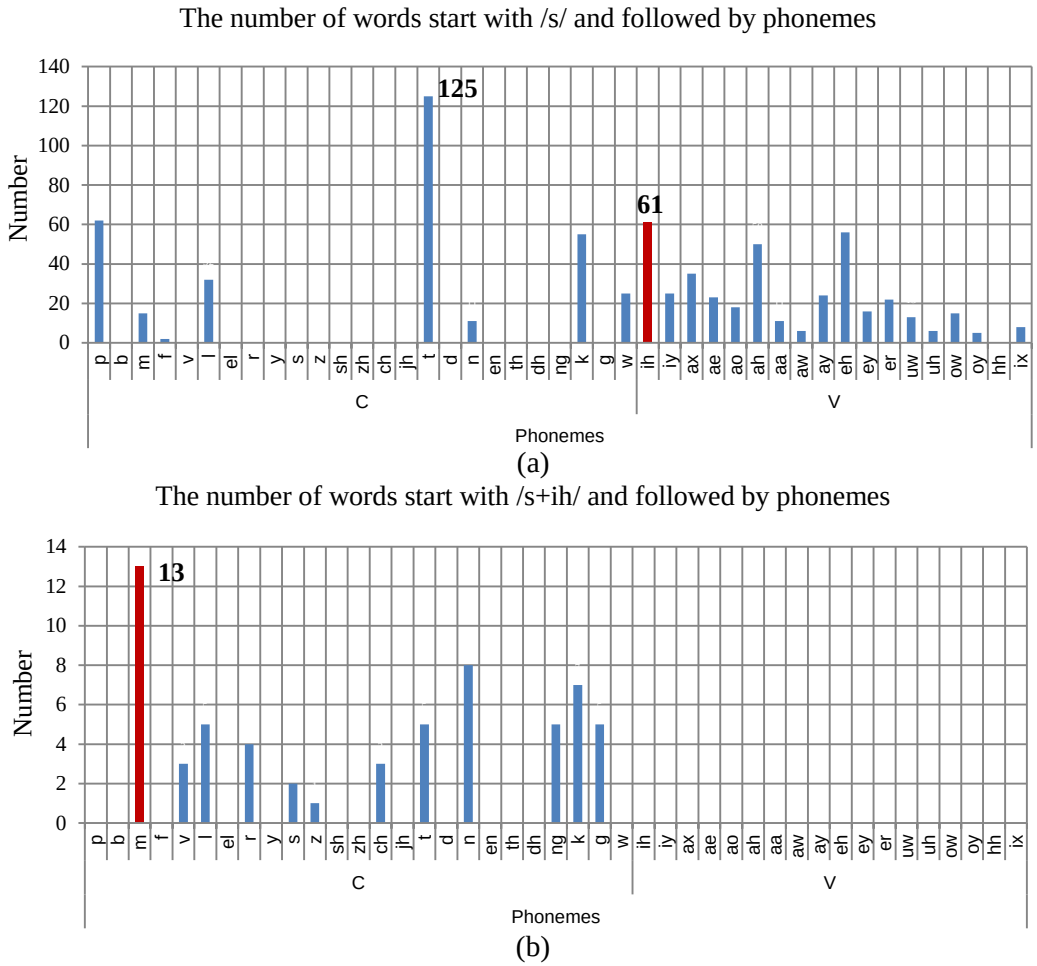


Figure 3-12: The phoneme population in TIMIT's transcribed corpus in the (a) second and the (b) third level of decision tree

This method of grouping words has a great benefit in a lip-reading system since searching for visual counterparts of words is limited to lower numbers of groups. This means, the searching process is not necessary for all phonemes and limited to 23 phonemes (Figure 3-12 (a)). In the third stage of searching, there is no need to look for any vowel (Figure 3-12 (b)).

3.4.5. BUILDING A LIMITED NUMBER OF WORDS DATABASE

The decision tree containing three levels would be obtained from the sequence of all the hierarchal combinations of phonemes in words. By tracing the decision tree, the maximum values of sequential combination of C and V in words are chosen as the target words set. Such combination includes the highest number of V or C in the beginning, the second and third position of words. Therefore, it is possible to claim that the extracted words from decision tree with the maximum occurrences of C and V

sequences are the most frequent words that are appearing in the TIMIT's vocabulary database. Therefore, it validates the processes of deriving the signatures for representing the visual words since the signatures have been derived for such systematically organized family of words.

The decision tree according to Table 3-3 and Table 3-4 can also demonstrate the phoneme sequential behaviour. By proceeding to the first level, 76.8% of words start with consonants and 23.2% with vowels. In the second level 21.4% of words start with consonants followed by consonants and 78% by vowels. In this level, 87.7% of words which start with vowels, proceed to consonants and 12.2% to vowels. The string of /CC/ in the second level is transferred to the third level where 8% of them proceed to consonants and 92% represent vowels in their third positions. The /CV/ sequences of words are transited to the third level where 95.6% of them have consonants in their third place and 0.44% of them have vowels. The sequence /VC/ detected in the second level is transferred to the third level where 56.7% of them proceed to consonants and 43.3% to vowels in their third positions. The /VV/ sequence of words transit to the third where there is no vowel in the third level.

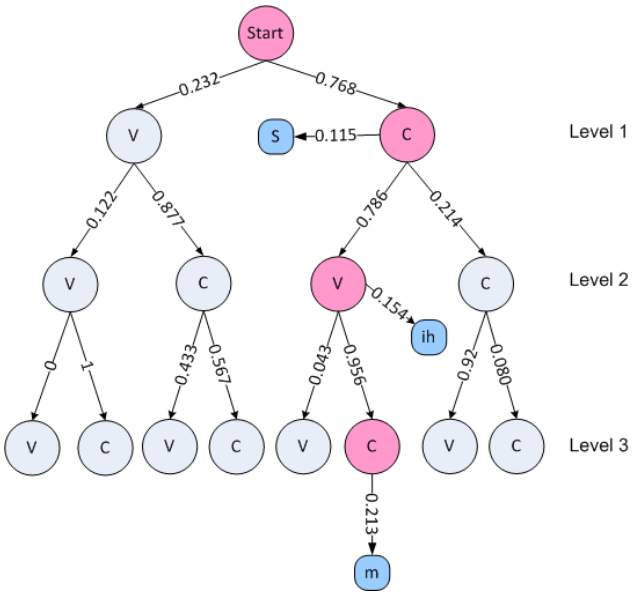


Figure 3-13: The hierarchical analysis of consonant structure of the TIMIT's transcribed corpus in form of decision tree for three levels

In Table 3-6, the thirteen selected words are analysed based on the phoneme symbols in different position of the words. The results are derived manually. All of them are

sharing the phoneme /s/ that occurs with maximum number (11.5%) in the first level. Similarly, in the second and third levels, the phonemes /ih/ (15.4%) and /m/ (21.3%) appears with maximum frequency. The shortest and longest words, which consist of four and thirteen phonemes, are ‘simmer’ and ‘sympathetically’.

Level														
1	2	3	4	5	6	7	8	9	10	11	12	13	Phonemeic representation	Words group
/s/	/lh/	/m/	/p/	/l/	/ax/	/t/	-	-	-	-	-	-	/s ih m p l ax s t/	simplest
						-	-	-	-	-	-	-	/s ih m p l ax/	simpler
					/iy/	-	-	-	-	-	-	-	/s ih m p l iy/	simply
				/el/	-	-	-	-	-	-	-	-	/s ih m p el/	simple
				/ax/	/th/	/eh/	/t/	/ih/	/k/	/ax/	/l/	/iy/	/s ih m p ax th eh t ih k ax l iy/	sympathetically
				/ow/	/z/	/iy/	/ax/	/m/	-	-	-	-	/s ih m p ow z iy ax m/	symposium
				/t/	/ax/	/m/	-	-	-	-	-	-	/s ih m p t ax m/	symptom
			/b/	/ax/	/l/	/ih/	/z/	/em/	-	-	-	-	/s ih m b ax l ih z em/	symbolism
						/ay/	/z/	-	-	-	-	-	/s ih m b ax l ay z/	symbolize
				/aa/	/l/	/ih/	/k/	-	-	-	-	-	/s ih m b aa l ih k/	symbolic
				/el/	/z/	-	-	-	-	-	-	-	/s ih m b el z/	symbols
			/ax/	-	-	-	-	-	-	-	-	-	/s ih m ax/	simmer
			/eh/	/n/	/t/	-	-	-	-	-	-	-	/s ih m eh n t/	cement

Table 3-6: The arrangement of the selected words and their phonemic representations in different levels of appearance

The benefit of suggested designing process for the corpus can be observed. For instance, in the level 9 of phonemes in Table 3-6, searching for the visual counterparts can be restricted to 3 phonemes /ih/, /m/ and /em/. This means there is no need to search visual clues of all phonemes. After level 9, the word ‘sympathetically’ will be chosen since there is no other word define after level 9.

3.5. SUMMARY

Extracting the visual information from a systematic sequence of phonemes enhances audio-visual speech recognition systems, increases the accuracy of lip-reading systems and the realism of visual speech synthesizers. In this chapter, a method for designing the corpus, which consists of a group of systematically related words, has been suggested.

In this work, it is desirable that the visual word follow a structure in its transcribed word domain. This new method is based on selecting a family of words which share specific phonemic rules among its members. A corpus, which contains English words and their phonemic representation, has been chosen. Based on the phoneme’s probability of appearance, a group of words has been selected from the corpus. The

group of words will be used for extracting the visual information in the next chapter. They are selected according to the maximum number of repeated consonants (C) and vowels (V) in the first three positions (triphones) of words in TIMIT's transcribed corpus.

The corpus in form of triphone /CVC…/ has been extracted from the decision tree by considering firstly the maximum occurrences of consonants and vowels and secondly the maximum occurrences of consonants and vowels in each branch for three levels in the decision tree.

The process has been facilitated by searching the consonants and vowels in the phonemic representations of the words. In the first level of search, the total number of consonants and vowels in the first position of words has been determined. In the second level same process has been performed but separately for the previously detected consonants and vowels. Similarly, the third level of searching determined the number of consonants and vowels in the first three sequential positions of words as /CCC…/, /CCV…/, /CVC…/, /CVV…/, /VCC…/, /VCV…/, /VVC…/, and /VVV…/. A decision tree geometrically visualized these three steps. The number of detected consonants and vowels in each level l_n is $2^n, n \in \mathbb{N}$. The decision tree like a map shows all possible paths to generate words. In this work, the desirable path has been defined based on the maximum numbers of a consonant or a vowel in each level. The maximum probability appeared in each stage has been the guideline for selection. In the next level, the maximum occurrences of C or V combinations would be selected. In this way, the coverage of words with similar phonemes has been maximized. Obviously, grouping all words takes longer than three stages.

A very important benefit of such approach is to create the logical and actual bimodality in visual and speech domains. It means the speech domain would be formulized in transcription level of words by their phonemic information. The other significant strength of structured phonemic analysis approach has been the inclusion of articulation effect. Furthermore, if a lip-reading system uses such corpus, it will not need to analyse all the visual counterparts of phonemes since the corpus limits the possible phonemes in each stage of searching. For future search algorithms which could be employed for recognition of visual clues, this information reduces the timing cost significantly.

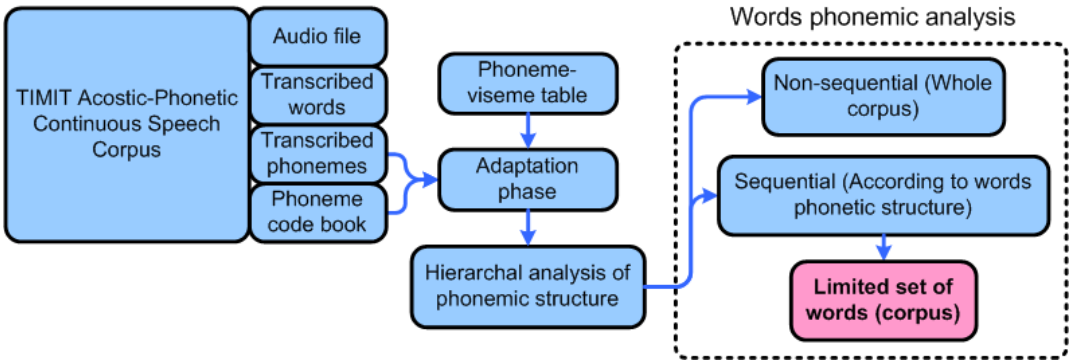


Figure 3-14: The designing steps toward extracting a limited set of words

In Figure 3-14 more details about the final proposed approach for obtaining the corpus are illustrated. The outcomes of data acquisition phase in the form of raw visual speech sample sets are extracted from subjects during articulation of a limited number of chosen words.

From such category of words, the visual feature data is extracted to form the visual speech signals. The key point is relating the information in the sequence of frames to words. The mathematical expressions and signatures processes are conducted on this set of words. In the next chapter, the framework of visual data acquisition from the group of words in terms of basic notations, definitions, processing and statistical analysis will be described.

Chapter 4

LIP CONTOUR PARAMETERIZATION AND VISUAL DATA ACQUISITION

The detailed characteristics of visual feature points on the lip in each frame and how it should be defined has been described in this chapter. As discussed in the literature review, all colour-based approaches for localizing the lip region are based on analysing the colour properties of skin and lip. For instance, in some cases it is difficult to separate the skin colour and lip colour even with by watching it. Obviously, such detection would be more difficult to implement on software level. Therefore, the robustness of image processing methods will affect the segmentation accuracy, although most of colour-based algorithms are automated. The other important factor in colour-based lip detection is the illumination effect. In this case, the environmental light will affect the natural Figure 4-1skin colour's appearance.

In this work, due to these issues for accurate lip detection and the necessity of precise formulation of lip's movements, the lip localization and feature points extraction are performed manually. Although the process is time consuming, the results are supervising and reliable for formulating the lip's movements. The region of interest (ROI), which all of information of lip's geometry is located in, is defined as an area on each frame.

The ROI is set to be a rectangular area around the speaker's mouth. In this work, the ROI is also manually adjusted to cover the lip region on each frame during video recording of the words' articulation. It means its size is adjusted according to to a safe boundary of lip during articulation in order to cover the maximum possible lips' dimensions.

4.1. ESTIMATING LIP GEOMETRY

The aim of this section is defining a visual sample set, which describes the lip geometry. The process is also referred to as parameterizing the lips. The visual samples are extracted from these visual feature points located on the upper, lower position and corner outer lip contour, denoted by superscript u, l and c, respectively. In terms of the pixel positions, the visual feature points are:

$$fp^{u}(i) = p({x_i}^{u}, {y_i}^{u}) \tag{4-1}$$

$$fp^{l}(i) = p({x_i}^{l}, {y_i}^{l}) \tag{4-2}$$

$$fp^{c}(i) = p({x_i}^{c}, {y_i}^{c}) \tag{4-3}$$

where $i = \{0, 1, 2, \dots, (F-1)\}$. The total number of frames, which covers the lip's movement is denoted by F. The frame by frame change of visual features creates the visual speech sample sets. In Figure 4-1, a typical relation of pixels and frame sequence is shown.

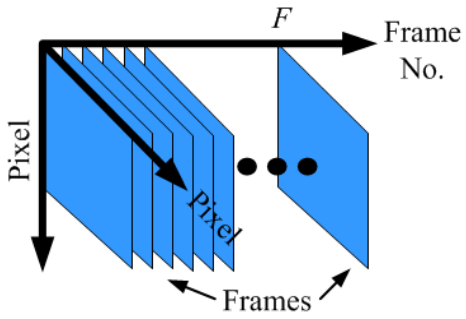


Figure 4-1: A typical representation of frame sequence

The distance of each feature point has been measured from the top and left boundaries of the ROI (x and y axis) and the desirable measurements of lip movements are depicted in Figure 4-2. The relations between the feature points described in the lip's geometry, define the visual speech samples. The feature points cannot be used directly. The measurement of feature points needs to represent the degree to which of lip opens or closes. The pink dashed lines show the 14th frame of the second test subject, during articulation of word "Symbol".

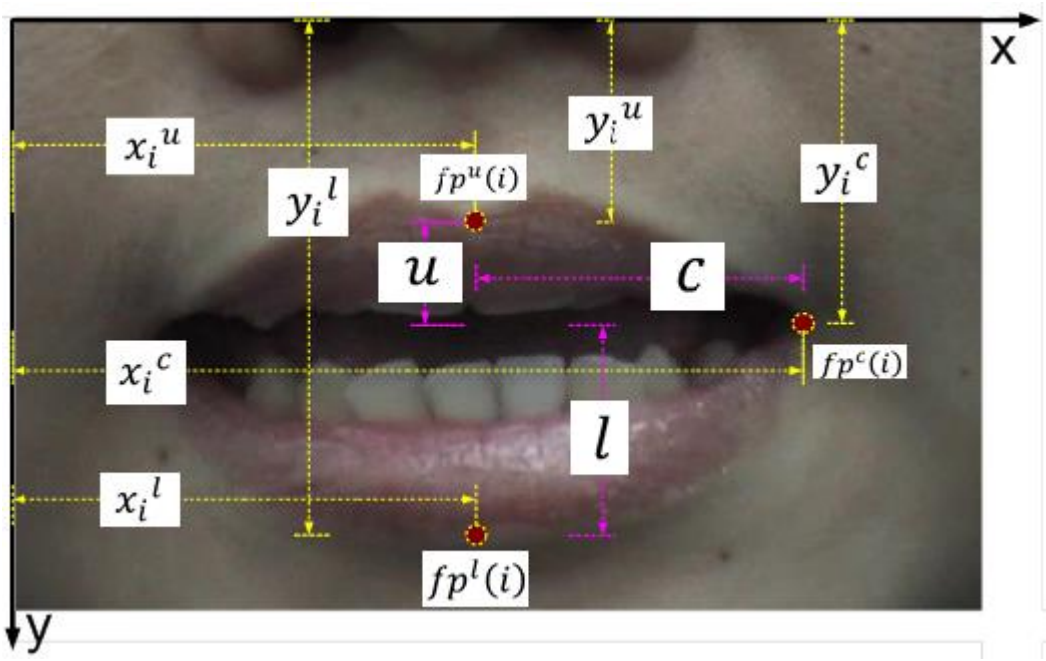


Figure 4-2: The geometrical representation of lip in a frame's ROI and visual features relations of pixels in Cartesian coordinates

The upper feature point is determined by subtracting $y_i{}^u$ from $y_i{}^c$. The distance between the lower feature point from the horizontal line connecting lip's corners would be described determined by subtracting $y_i{}^c$ from $y_i{}^l$. These parameters show how much lips are opened in vertical directions. Similarly, the distance of corner feature point from the vertical line which connects the upper and lower feature points would be determined by subtracting $x_i{}^u$ from $x_i{}^c$. If the upper and lower feature points pixels are collinear, then $x_i{}^u$ will be interchangeable with $x_i{}^l$. Due to the jaw movement, unwanted fluctuations may appear more in lower feature point during articulation than the upper feature point since unlike the upper lip which is fixed to the mandible, the lower lip is controlled by the jaw. Therefore, $x_i{}^u$ represents extracted pixel values of the corner visual vector. Applying such constriction to Eq. (4-1) to (4-3) leads to:

$$fp^u(i) = P(x_i{}^u, y_i{}^c - y_i{}^u) \tag{4-4}$$

$$fp^l(i) = P(x_i{}^l, y_i{}^l - y_i{}^c) \tag{4-5}$$

$$fp^c(i) = P(x_i{}^c - x_i{}^u, y_i{}^c) \tag{4-6}$$

The pixel differences are noted by $u = \{u_i\}$, $l = \{l_i\}$ and $c = \{c_i\}$, $i = \{0, 1, 2, \dots, F - 1\}$ forming the visual speech sample sets for the upper, lower, and corner feature points sample sets $FP^u(i)$, $FP^l(i)$ and $FP^c(i)$. These vectors now are measuring the feature points according to the Cartesian coordinate system that has been defined within the lips geometry. Therefore, they can be represented by:

$$FP^u(i) = \{u_0, u_1, u_2, \dots, u_{F-1}\} \tag{4-7}$$

$$FP^l(i) = \{-l_0, -l_1, -l_2, \dots, -l_{F-1}\} \tag{4-8}$$

$$FP^c(i) = \{c_0, c_1, c_2, \dots, c_{F-1}\} \tag{4-9}$$

where i is the frame sequence and $i = \{0, 1, 2, \dots, F - 1\}, m = 1, 2, 3, \dots, w$. These vectors can now describe the measurements of the lip during uttering words. The feature points' sample sets are described by W_m and can be represented in matrix form as:

$$W_m = \begin{bmatrix} FP^u_{W_m}(i) \\ FP^l_{W_m}(i) \\ FP^c_{W_m}(i) \end{bmatrix} \tag{4-10}$$

So far the database consists of uttering one version of the words by one speaker. This is not usually the case and consequently the speaker is asked to repeat the words several times. Each word W_m has been generated by p speakers. In this work, the number of speakers is two. Based on that, a visual sample set ${W_m}^{SP}$ has been gathered from speakers (subjects) $SP = \{1, 2, 3, \dots, p\}, p \in \mathbb{N}$. The video files database VW can be represented by:

$$VW = \{W_1, W_2, W_3, \dots, W_m\} \tag{4-11}$$

The total versions of words VW produced by all speakers SP are denoted by VWS:

$$VWS = \{{W_m}^{SP}\} \qquad m = \{1, 2, 3, \dots, w\} \qquad SP = \{1, 2, 3, \dots, p\} \tag{4-12}$$

Each word has been repeated r times by the speakers. Therefore, a modification applies to subscript m in ${W_m}^{SP}$ that is denoted as ${W_{m_q}}^{SP}, q = \{1, 2, 3, \dots, r\}$. The total

versions of words VW produced by all speakers SP with repetition r would be shown by $VWSS$:

$$VWSS = [{W_{m_q}}^{SP}] \tag{4-13}$$

where $m = \{1, 2, 3, \dots, 13\}$, $SP = \{1, 2\}$ and $q = \{1, 2, 3, \dots, 20\}$. In order to increase the accuracy of articulation and to maintain the practicality of experiments, the maximum value of q has been set to 20. The total number of pronounced words video files $VWSS$ is $m \times p \times q$. In this stage, it is possible to include the visual speech vectors in definition of ${W_{m_q}}^{SP}$ words. For obtaining a representations of the visual speech sample sets, Eq. (4-7) to (4-9) have been modified to describe a single expression for the visual data as follow:

$$FP^{u}_{W_m}(i) = \{u_0^{W_m}, u_1^{W_m}, u_2^{W_m}, \dots, u_{F_m}^{W_m}\} \tag{4-14}$$

$$FP^{l}_{W_m}(i) = \{l_0^{W_m}, l_1^{W_m}, l_2^{W_m}, \dots, l_{F_m}^{W_m}\} \tag{4-15}$$

$$FP^{c}_{W_m}(i) = \{l_0^{W_m}, l_1^{W_m}, l_2^{W_m}, \dots, l_{F_m}^{W_m}\} \tag{4-16}$$

It should be mentioned that in Eq. (4-8), the absolute values of samples are taken as the visual speech samples since the amplitudes of them are the subject of studying not their directions. In the next section, these sample sets are interpreted as a sequence of samples. The lip's dimension varies within a boundary, which also defines constraints on the range of feature point's sample set, during articulation. The boundaries of feature points are determined based on the ROI. Therefore, the minimum and maximum closures of lips during uttering a particular word are:

$$Min^{u}_{W_{m_q}} < FP^{u}_{W_m}(i) < Max^{u}_{W_{m_q}} \tag{4-17}$$

$$Min^{l}_{W_{m_q}} < FP^{l}_{W_m}(i) < Max^{l}_{W_{m_q}} \tag{4-18}$$

$$Min^{c}_{W_{m_q}} < FP^{c}_{W_m}(i) < Max^{c}_{W_{m_q}} \tag{4-19}$$

4.2. VISUAL DATA ACQUISITION

Obtaining the ideal pronunciation of spoken words and extracting visual data is challenging in visual data acquisition. In this section, methods of selecting and extracting a fixed length of frames, including the visual feature data of each word, are

suggested. The following methods enhance the performance of articulation acts by speakers. It is possible to represent the transcriptions of detected words to the speakers and ask them for uttering, but since the proper pronunciation should be optimized, broadcasting words' audio files (Bowyer, 2006-2013) associates more accuracy in the manner of articulation.

In this work, the other usage of audio representation is predicting the number of frames in visual speech signal data by adding the anticipation effect. In other words, the length of a visual speech sample set has been approximated from the length of audio data. This can be useful as an index to choosing the most similar frame lengths between the speaker's visual data and the reference audio data. Although, in case there is no matching between predicted number of frames and speakers pronunciation of words, averaging method would be applied for choosing the words video files from speakers' versions of words.

The subjects used in this study are two female non-native English speakers. They are aged 26 and 33 years old with no hearing impairment or articulation disabilities. Obviously, the ability of pronouncing the words is related to being native in English language. For example, German native speakers find the phoneme 'w' difficult to pronounce and in attempt to alleviate this problem prior to the text the audio file would be played to the subjects in order to minimize such difficulties. In order to select the visual data, manual editions of frames are conducted with the aid of approximated frame length that anticipation effect has been added to them from audio files broadcasted to the speakers. However, since the aim of this thesis is to verify the methodology, the differences in pronunciations do not significantly affect this work.

The other possible approach for obtaining more ideal pronunciations is gathering more samples from each word. Therefore, the subjects had been asked to repeat each words several times.

The accuracy of extracting video frames will be achieved by considering the audio signal lengths and adding anticipations starting with selecting appropriate recorded video file. In the next stage, the video frames will be extracted. Afterwards, the extracted 40 versions of each word (2 speakers articulate 20 times) will be compared with ideal video lengths for choosing as the best match. In case of mismatching, an

alternative method will be used by finding the trimmed mean among 40 versions of articulated words. Based on the similarity to trimmed mean values, the suitable video frames will be selected.

In the next phase, the visual speech samples will be extracted automatically for comparing with manual extraction. The comparison justifies the usage of manual extractions of visual speech sample sets.

4.2.1. UNIFYING THE FRAME SIZE

Each sample set may have different number of frames to the others. For creating speaker independent (SI) visual words signature expressions, the speakers' visual data has been analysed in order to unify the several visual data for each word. The selected signature expressions inherit common visual words from the speakers. It means the visual words' signatures could be a combination of visual signals provided by the speakers based on particular criteria.

In other words, articulation of the same word by the same speaker two times will result in different frame lengths. In this work, each speaker generates 20 versions of the visual data for one word to increase the accuracy of words articulation. Therefore, the 40 versions for each word would probably have different number of frames. It is desirable that all of these samples converge to a common number of frames.

On the other hand, there are other factors, which interfere reaching a common number of frames. For instance, the method of selecting the frames that contains the lip's motion information in its beginning and ending is a challenging task. In addition, the manual edition in selecting frames has been employed and tried to be conducted carefully. These two issues cause variation in pronounced versions of a word. In the following sections, two methods for optimizing the number of frames in each word are suggested. The output of data preparation and acquisition would consider variations of samples taken from feature points. These feature points sets are called visual speech vectors.

In the visual data acquisition phase, the visual information of the transcribed words in video file format has been extracted. There are some factors, which could reduce the performance of speakers articulation act, and in parallel there are other factors that

enable increasing it. To enhance the accuracy of speakers' performance, playing the words' audio files for speakers has been associated to the visual data acquisition phase and the schematic of process is shown in Figure 4-3.

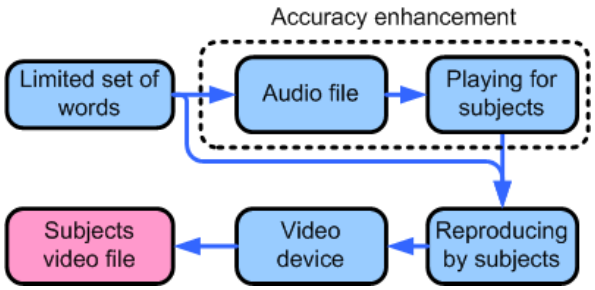


Figure 4-3: The suggested process of visual data acquisition

The subjects had been asked for clear and complete pronunciation of words (lip closure before and after pronunciations) with neutral facial expressions. The sessions have been conducted under normal lighting conditions and no marking has been used on the faces. The camera properties are:

- Name: Sony DCR-SR40 30 GB Handycam Camcorder
- Movie mode: MPEG2
- Resolution: VGA
- Video pixels: $340^{kpixels}$
- Recording format: Dolby® Digital AC3 (2 channels)
- Frame rate: 30^{fps}

4.2.2. ACCURACY ENHANCEMENT OF VISUAL DATA

The methods of obtaining the approximated number of frames, which fully is covered in uttering the words, are described in this section. To achieve accurate selection of extracted visual sample sets from all 40 versions of each articulated word, the timing information of speech signals in the reference audio files have been studied. On the other words, the audio signal timing provides which frames should be selected for starting and stopping the articulation. The number of samples has been calculated in the audio files and their time durations have been measured in seconds. Based on the audio files timing information, the approximated frame numbers, which are appeared in the video files, are calculated by multiplying audible words timings and the recording device frame rate per second. The number of frames involved in calculating

the time can be simply determined by $F_{W_m} = S_{W_m} F$ where F_{W_m}, S_{W_m} and F are visual data's frame length, word's duration in second and recording device frame rate, respectively.

The data has been extracted manually, frame by frame. The video file has been reviewed for detecting proper samples in this manner to ensure that the subjects have pronounced the words completely; i.e., the mouth should not remain opened between articulating the words. The other issue is the shape of mouth before pronunciation. It should not be opened long before speaking. Therefore, some sample sets would not appropriate candidates for processing and have to be discarded. The key frames corresponding to visual appearance of words pronunciation has been extracted in this way. Therefore, face and mouth detection methods i.e. colour segmentation or AAM have not been used in this work.

For segmentation of each uttered word, the 'Windows Movie Maker software' has been used. The acceptable video format for this software is WMV format. Since the recorded video formats of the lips movements are in MPEG2, the video files are converted to WMV formats using video converter (Leawo, 2012) by 'FREESTUDIO' software. It should be mentioned that the video frames are converted from RGB to grayscale colour map since the number of colour spaces reduces from three to one.

The three feature points are located and marked using MATLAB 'datatip' option on each frame. Each pixel's value needs to be transformed into the Cartesian space. The pixel's x and y values are the distances from upper horizontal boundary and left vertical boundary respectively in each image, as it described in section 4.1. In Figure 4-4the frame sequences of articulating the word 'simmer' by one of the speakers has been represented.

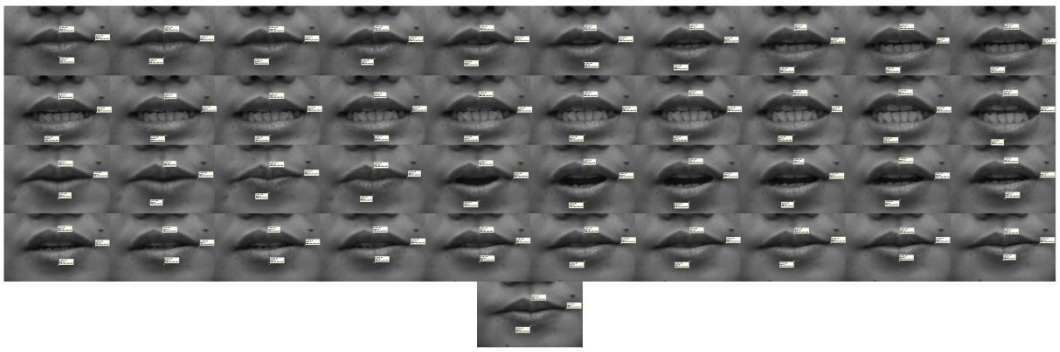


Figure 4-4: The frame sequence in articulation of the word ‘Simmer’ in grayscale

- **ADDING ANTICIPATION**

Each word would have basic number of frames related to its audio signal duration. These numbers of frames are exactly included during articulation while in visual domain the mouth shaping starts before and lasts longer than the speech signal. Such delay between audio and visual lip movement is called anticipation. In Figure 4-5, the effect of anticipation is represented by comparing a speech signal against corresponding video frames. The video frames are categorized into two regions according to the presence of speech signals in common and anticipated periods. In common period, the audio signal has fully overlapped by video frames. The anticipation effect in visual domain occurs in soundless configuration of mouth geometry as lips are intended to generate speech signals during rest state. In the end of generating speech signal reaches the rest state.

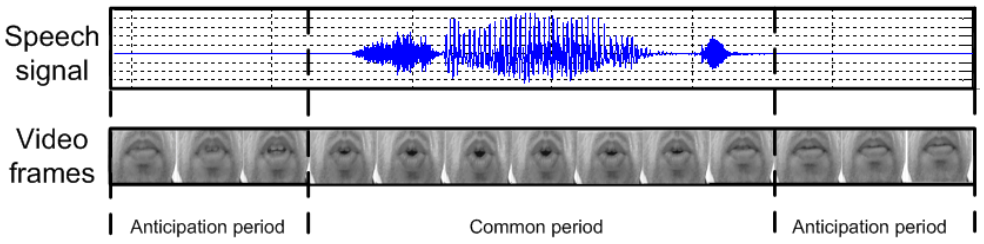


Figure 4-5: Anticipation effect

Applying the anticipation effect is completely dependent on the priori knowledge about the visual data. This phenomenon clearly appears before and after silent articulation in visual domain. This quality cannot be supported by a quantity. Therefore, the anticipation margin is kept by adding extra three frames in the beginning and ending of words frame sets. These frames cover the smooth

transformation of visual speech sample sets from the rest position of lip to the point where audible speech is generated. Similarly, after audible speech is stopped, these frames cover the transition of lip to its rest position. The audio samples are extracted using the 'MATLAB' software. The audio time durations (44.1^{kbps}) has been determined dividing by sampling frequency of the audio files.

In Table 4-1, the details of speech signal duration, the approximated frame length, and the approximated frame length after adding anticipation effect (3 frames in the beginning and in the end) are demonstrated. The use of 3 frames is not fixed and could vary based on different editors. The experimental selection of frames and the approximated speech signals durations from captured video files corresponding to each word has to be demonstrated. In this stage, the word 'simply' has been matched to the approximated frame numbers from the audio file.

The audio signals of each word played for speakers' articulation enhancement has been represented in Figure 4-6. The audio files are recorded from the Cambridge online dictionaries (Press, 2013). The audio duration of words are also used for observing the match between speakers' performance as well as manual selection of the frames.

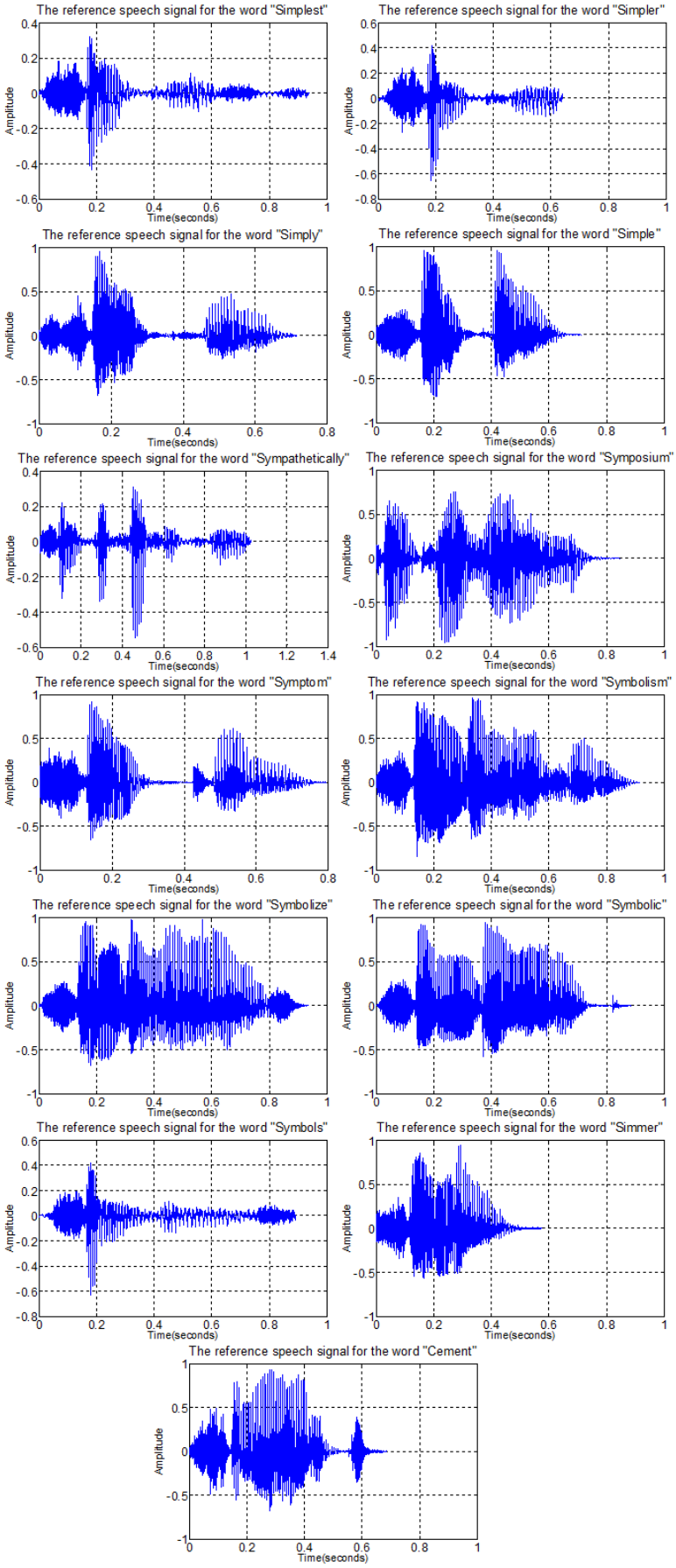


Figure 4-6: The speech waveforms corresponding to the chosen text corpus

Table 4-1: The number of frames after applying anticipation effect to each word

		Audio file analysis		
		Speech duration S_{W_m} (second)	Approximated frames lengths covering the speech signal (fps)	Approximated frames lengths (fps) after adding anticipation effect
Words	Simplest	0.939	29	35
	Simpler	0.640	20	26
	Simply	0.717	22	28
	Simple	0.711	22	28
	Sympathetically	1.025	31	37
	Symposium	0.846	25	31
	Symptom	0.794	24	30
	Symbolism	0.915	28	34
	Symbolize	0.937	29	35
	Symbolic	0.884	27	33
	Symbols	0.890	25	31
	Simmer	0.578	18	24
	Cement	0.685	21	27

The frame sequences contain visual data that has been selected according to the best-matched speakers' visual speech frame lengths and the anticipated words' frames lengths. This process is shown in Figure 4-7.

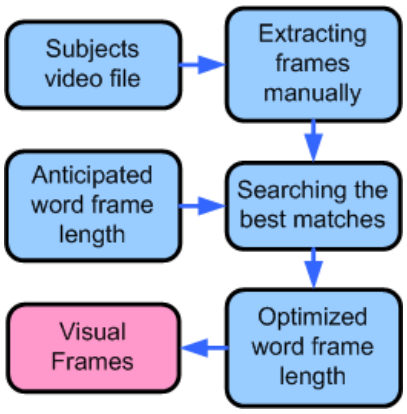


Figure 4-7: The suggested method to increase the amount of similarity between the audio source file and the selected visual data

The most coherent pronounced visual data by speaker in term of frame number $F_{W_{m_q}SP}, m = \{1, 2, 3, \dots, w\}$ needs to be found among the total 40 repeated words versions. In order to do so, the index of choosing one video file out of 20 repeated

versions of each word by Speaker 1 and Speaker 2 is the nearest $F_{W_{m_q}{}^{SP}}$ to the approximated number of frames from audio files.

4.3. ALTERNATIVE SEARCHING FOR THE BEST MATCHES

The anticipated frames, which are covering the audio file, may not be matched to the extracted visual data frame numbers. Therefore, the visual data was extracted using trimmed mean function and rounding them to the nearest integers as the alternative model as it is shown in Figure 4-8. It should be mentioned that, the visual data of both speakers is the subject for calculating the trimmed mean.

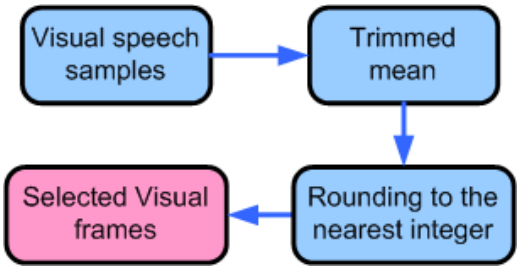


Figure 4-8: Alternative method for selecting visual speech sample sets

In case of detecting more than one visual data with the same value of trimmed mean, the final visual speech sample sets, which would be rounded to the nearest integers and the average of the extracted visual speech sample sets would be chosen.

In Table 4-2, the lengths of the video files $F_{W_{m_q}{}^{SP}}$ are represented. By comparing the approximated frame lengths from the audio files in Table 4-1 and Table 4-2 it can be realized that no matches exist. Therefore, the alternative method that uses trimmed means of frame lengths has been used to select the video files. The video files for the Speaker 2 are deleted during visual data extraction process since they are not accurate enough to use.

The final decision on the lengths of video files for each word has been selected as the rounded trimmed mean to the nearest integer and it has been shown in Table 4-3.

Table 4-2: The speakers video frame lengths

		Words frame lengths $F_{W_{m_q}}$												
	q	Simplest	Simpler	Simply	Simple	Sympathetically	Symposium	Symptom	Symbolism	Symbolize	Symbolic	Symbols	Simmer	Cement
Speaker 1	1	58	49	56	43	78	47	47	45	58	56	47	39	39
	2	54	51	54	48	57	47	48	47	56	52	50	44	46
	3	52	53	63	53	66	45	44	48	60	47	46	49	39
	4	56	50	53	53	64	40	43	48	59	51	49	43	47
	5	48	46	53	50	65	43	51	45	53	50	58	47	45
	6	52	48	58	45	63	44	48	46	56	54	49	52	39
	7	52	49	47	43	66	42	43	48	53	50	41	40	41
	8	56	49	50	44	60	43	44	50	54	47	48	48	45
	9	55	47	66	46	64	42	50	51	52	51	47	44	39
	10	51	49	48	45	65	41	47	44	59	52	48	47	42
	11	54	49	52	43	58	43	49	41	54	56	43	41	38
	12	58	46	58	46	61	42	55	48	57	54	43	40	46
	13	56	47	56	43	63	43	51	50	55	57	47	41	40
	14	59	44	49	48	62	45	45	47	58	52	43	44	43
	15	55	44	51	51	64	43	54	44	53	45	48	43	42
	16	51	49	53	47	59	45	40	45	58	50	41	47	41
	17	49	41	52	46	68	42	50	45	52	57	47	44	38
	18	57	53	40	52	62	45	43	42	62	58	48	45	41
	19	51	50	52	51	61	46	45	50	60	55	47	45	45
	20	65	49	55	56	70	53	78	49	64	56	49	47	45
Speaker 2	1	55	55	43	41	57		44	65	52	46	47	38	43
	2	51	58	40	43	58		46	45	49	41	51	44	45
	3	48	57	51	52	58		43	59	47	48	48	39	46
	4	49	50	52	43	58		44	49	48	47	48	44	50
	5	52	56	59	46	52		51	56	52	57	54	37	47
	6	51	50	50	53	61		41	62	47	44	52	35	48
	7	55	49	54	44	59		41	69	58	49	51	39	51
	8	59	48	54	46	59		44	55	51	50	52	43	48
	9	52	51	39	43	59		39	64	57	43	51	48	52
	10	56	51	40	43	56		42	70	54	48	50	41	46
	11	55	48	39	46	58		43	64	52	54	52	35	54
	12	57	49	59	44	60		38	58	59	55	55	40	49
	13	57	53	53	41	59		43	52	58	56	57	38	45
	14	56	54	40	43	53		44	62	47	62	52	40	44
	15	53	62	47	47	54		37	60	53	59	54	37	47
	16	59	49	42	43	53		42	63	51	57	50	36	42
	17	55	51	46	47	54		37	57	50	59	56	34	43
	18	50	46	47	47	50		41	66	50	56	54	38	48
	19	55	56	49	47	56		42	60	47	54	53	39	44
	20	55	48	43	45	56		47	64	51	53	56	39	54

Table 4-3: Alternative approach for calculating the visual speech sample sets lengths (number of samples)

		Trimmed mean	Rounded Trimmed mean
Words	Simplest	54.167	54
	Simpler	49.800	49
	Simply	50.333	50
	Simple	45.967	46
	Sympathetically	59.867	60
	Symposium	43.750	44
	Symptom	44.800	45
	Symbolic	52.367	52
	Symbolism	54.167	54
	Symbolize	52.467	52
	Symbols	49.633	50
	Simmer	41.267	41
	Cement	44.433	44

There are numbers of frame sets (highlighted by green colour in Table 4-2), which shared equal number of frames (Table 4-4). This means the numbers of frames are equal to two or more video frame sets. The amplitudes of these sample sets are averaged to obtain the final sample sets. In this situation, extracted visual sample sets in sharing video files will be the average of the amplitudes that are rounded to the nearest integers. These values are the length of words F_{W_m} in visual domain.

Table 4-4: The number of matched sample sets after calculating the trimmed means of visual speech sample sets lengths

		Data sets with equal frame length
Words	Simplest	2
	Simpler	4
	Simply	2
	Simple	2
	Sympathetically	2
	Symposium	-
	Symptom	2
	Symbolic	-
	Symbolism	3
	Symbolize	3
	Symbols	3
	Simmer	3
	Cement	2

Table 4-5: The approximated time duration of speech signals from selected video frames

		Experimental results	
		The frame length F_{W_m}	The approximation of speech signal time duration $t_i{}^f$ (second)
Words	Simplest	54	1.80
	Simpler	50	1.67
	Simply	48	1.60
	Simple	48	1.60
	Sympathetically	60	2.00
	Symposium	44	1.47
	Symptom	45	1.50
	Symbolic	52	1.73
	Symbolism	52	1.73
	Symbolize	54	1.80
	Symbols	50	1.67
	Simmer	41	1.37
	Cement	44	1.47
	Total length	642	21.40

By comparing Table 4-1 and Table 4-5, the total number of frames in experimental results is twice the approximate results. The total length of approximated speech signal in experimental results defers vastly from the audio files, which have been played for speakers. It implies the imperfection of articulated words in the visual domain. Apparently, the subjects' articulation times are twice the reference audio file. This fact implies that, on average, the speakers articulated the word slower than the reference audio file.

4.4. VISUAL DATA EXTRACTION

In this section, two approaches of visual speech samples extraction from the visual features will be described. These two methods are automatic and manual extractions. It will be shown that the accuracy of extracted visual sample sets via automatic method is lower than manual extraction. In order to achieve a higher accuracy of visual speech signals, the manual extraction of visual speech samples have been used in this thesis. Clearly, the automatic approach can be employed for possible future studies to reduce the computational time. However, for the future usage in automatic systems, the applicant should be aware of the amount of information loss (lip dimension) that can be tolerated after automatic extraction.

4.4.1. AUTOMATIC EXTRACTION

The automatic visual speech sample extraction in this work is inspired by lip segmentation part in the method proposed by (Wark & Sridharan, 1998). This method segments out the lip area from the background skin using pixel colour analysis. In the first step, images are separated to red, green and blue colour components. The lip area is determined by calculating the ratio of red to green components and finding upper and lower limits to the ratio. Due to imperfection of segmented lip area, the authors applied morphological closure to the segmented lip. Finally, edge detection algorithm is used for obtaining the outer lip boundary.

In this work, the visual speech samples are extracted according to block diagram in Figure 4-9, using MATLAB software. In order to segment the lip area from video files, the arithmetic relations of images colour components are studied. The selected video frames for each word contains the lip area. In order to reduce the computational cost, this area is reduced to cover the maximum opening of mouth with safe rectangular boundaries by using image cropping. Afterward, the pixel colour analysis is applied and the outer lip boundaries based on colour components are extracted. After studying the colour components, the most suitable level is chosen. In the next level, the colour image is transferred to binary image. In order to increase the accuracy of lip contour (reducing discontinuity of pixels), image dilation and erosion are applied. In the final stage, the maximum values on the upper, lower and two

corners of lip contour would be selected by geometrical analysis. More details about the process are described in the following paragraphs.

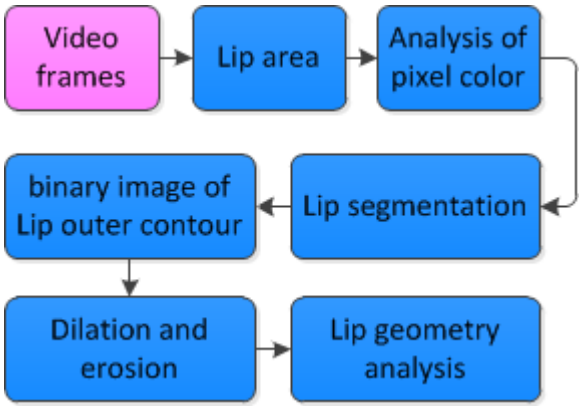
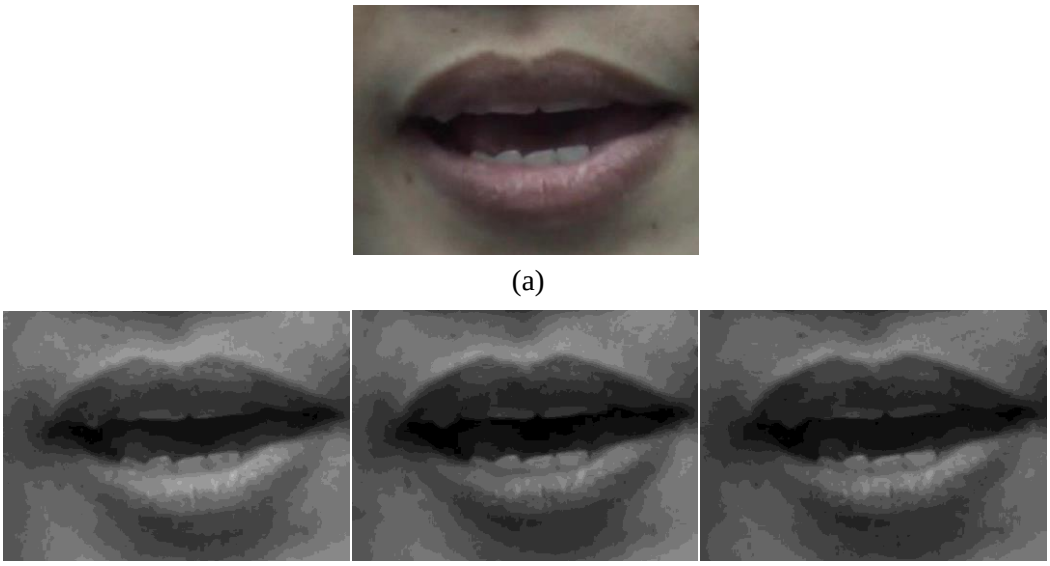


Figure 4-9: Automatic visual feature extraction

This method uses pixel colour properties of ROI to segment the lip's outer boundary. In the first stage, the RGB image of lip separated to red, green, and blue colour components as shown in Figure 4-10.



(a)

(b) (c) (d)

Figure 4-10: The 27th frame during articulation of word 'Cement' in (a) RGB, (b) red, (c) green and (d) blue colour component

In the next level, the arithmetic relations between colour components, e.g. $red - green$ are calculated. The relations of colour components are represented in Table 4-6. Visualizing colour component relations have been conducted using MATLAB command 'contour'. It simplifies the colours in an image by allocating integer values.

The results show the different interpretations of lip and skin colours as shown in Figure 4-11.

Table 4-6: The suggested arithmetic relations of colour components

Red	$Green$	$Blue$
$Red - Green$	$Red - Blue$	$Green - Red$
$Green - Blue$	$Blue - Red$	$Blue - Green$
$\frac{Red}{Green}$	$\frac{Red}{Blue}$	$\frac{Green}{Blue}$
$\frac{Green}{Red}$	$\frac{Blue}{Red}$	$\frac{Blue}{Green}$
$Red + Green$	$Green + Blue$	$Red + Blue$
$\frac{Red + Green}{Red}$	$\frac{Red + Green}{Green}$	$\frac{Red + Green}{Blue}$
$\frac{Red + Blue}{Red}$	$\frac{Red + Blue}{Green}$	$\frac{Red + Blue}{Blue}$
$\frac{Green + Blue}{Red}$	$\frac{Green + Blue}{Green}$	$\frac{Green + Blue}{Blue}$
$\frac{Red - Green}{Red}$	$\frac{Red - Green}{Green}$	$\frac{Red - Green}{Blue}$
$\frac{Blue - Green}{Green}$	$\frac{Red - Blue}{Green}$	$\frac{Red - Blueen}{BLue}$

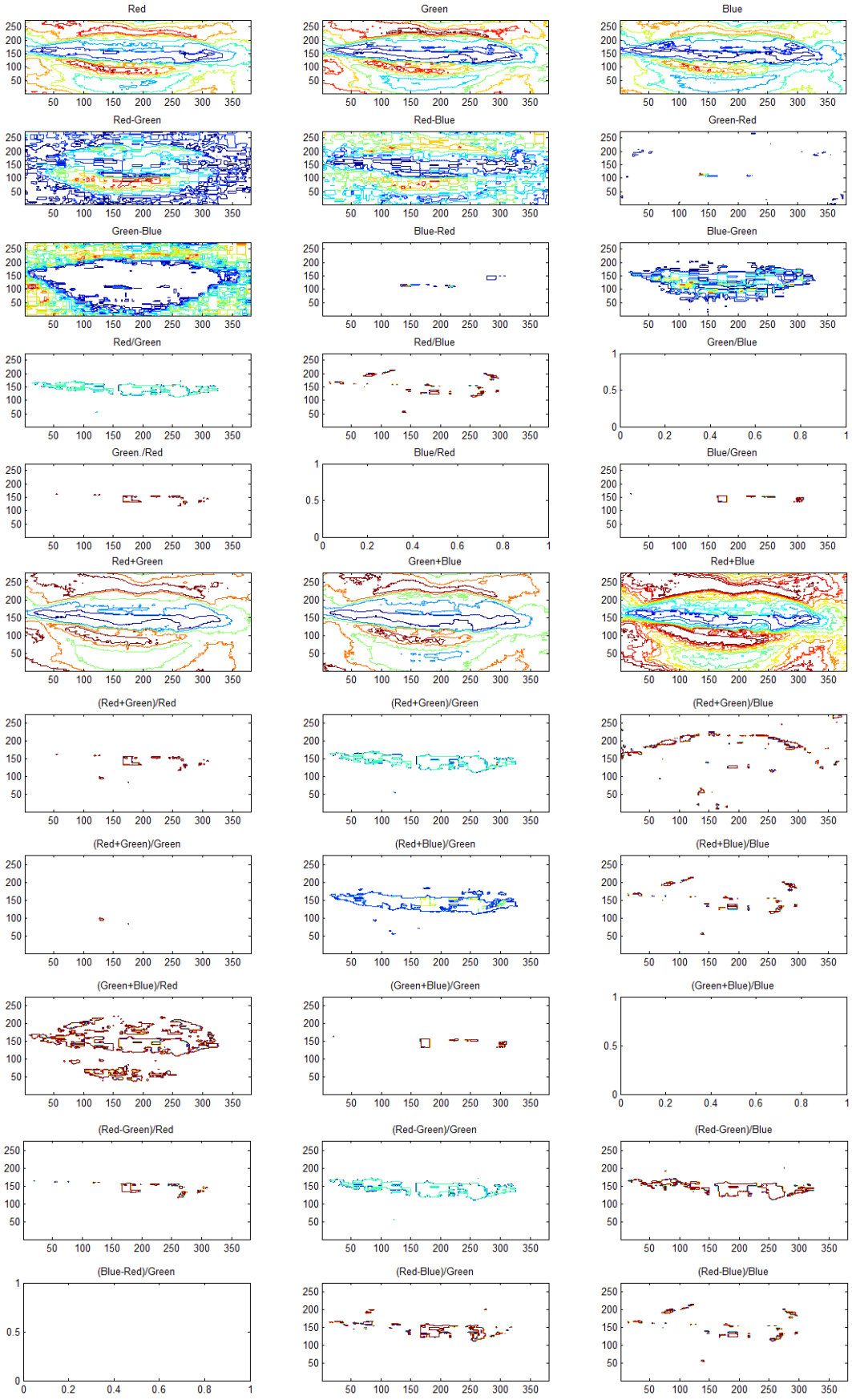
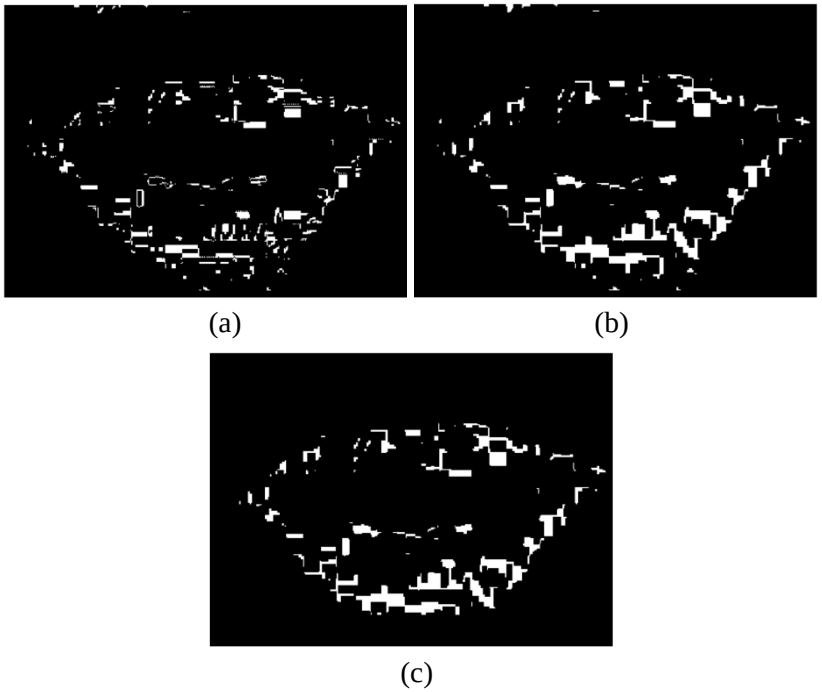


Figure 4-11: The colour analysis of the RGB planes

In the next level, by transferring these levels into binary images, the appropriate colour level can be chosen. Since the binary detected lip area does not cover the lip boundary completely (Figure 4-12 (a)), two image processing approaches called image dilation and erosion are employed to grow the lip area (Figure 4-12 (b)).

After choosing the colour level, it is possible to detect it in other areas other than lip region. It has been observed that the unwanted pixel colours mostly occur in the edges of ROI rectangle. Therefore, they are cleaned by setting rectangle edge to zero for certain lengths. The result is shown in Figure 4-12 (c).



(a) (b)

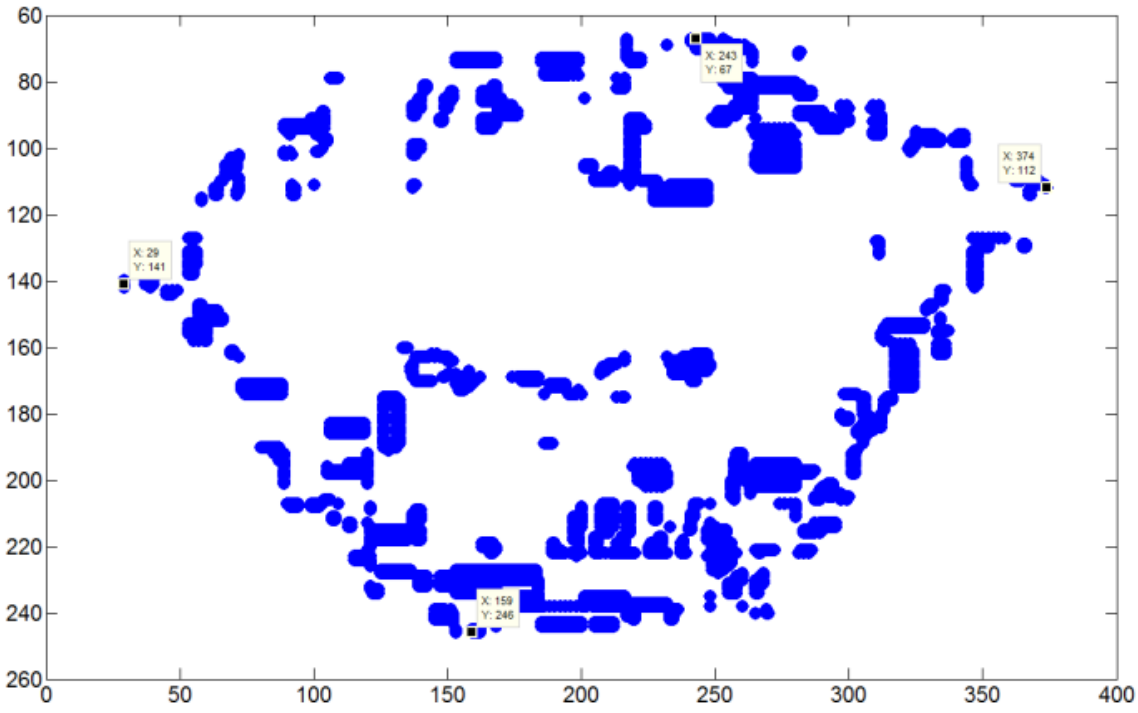
(c)

Figure 4-12: Lip boundary segmentation, (a) binary image of segmented Green-Blue plane, (b) after applying dilation and erosion, (c) the final lip area after binary cleaning

From Figure 4-12 (c), the rows and columns corresponding to logic 1 has been extracted. In the next phase, the minimum and maximum values across rows and column would be determined. The minimum and maximum of rows associate the y coordinate of upper and lower lip while the minimum and maximum of columns determines the x coordinate of left and right corner of the lip.

Finally, by referring to the row and column values, the x coordinates for upper and lower lip as well as y coordinate of corner lip will be determined. In Figure 4-13, the upper feature point is represented as $fp^u(27) = p(243,67)$, lower feature point is represented as $fp^l(27) = p(59,246)$, right corner feature point is represented as

$fp^c(27) = p(29,141)$ and left corner feature point is represented as $fp^c(27) = (374,112)$. In this work the left corner point is chosen for comparison and processing.



(a)

(b)

Figure 4-13: The detected feature points (a) after processing and (b) on the RGB image

One of the advantages of the suggested lip detection method is using colour components analysis. It can be useful to have a comprehensive sight about the ROI before further processing. In addition, during the study, it has been found that there is no need to use edge detection algorithms.

4.4.2. MANUAL EXTRACTION

The visual speech sample sets $FP^{u}_{W_m}(i)$, $FP^{l}_{W_m}(i)$ and $FP^{c}_{W_m}(i)$ have been extracted from the upper $fp^{u}(t_i)$, lower $fp^{l}(t_i)$ and the corner $fp^{c}(t_i)$ visual feature points on speakers' lips demonstrated in Figure 4-14 (a), (b) and (c), respectively for $i = \{0, 1, 2, \dots, F_{W_m} - 1\}$. At the first glance, the visual sample sets suffer from two issues. The first issue is the dissimilarities between starting samples and the second issue is the misbalance between the first and last samples. These issues will be addressed and solved in the next chapter.

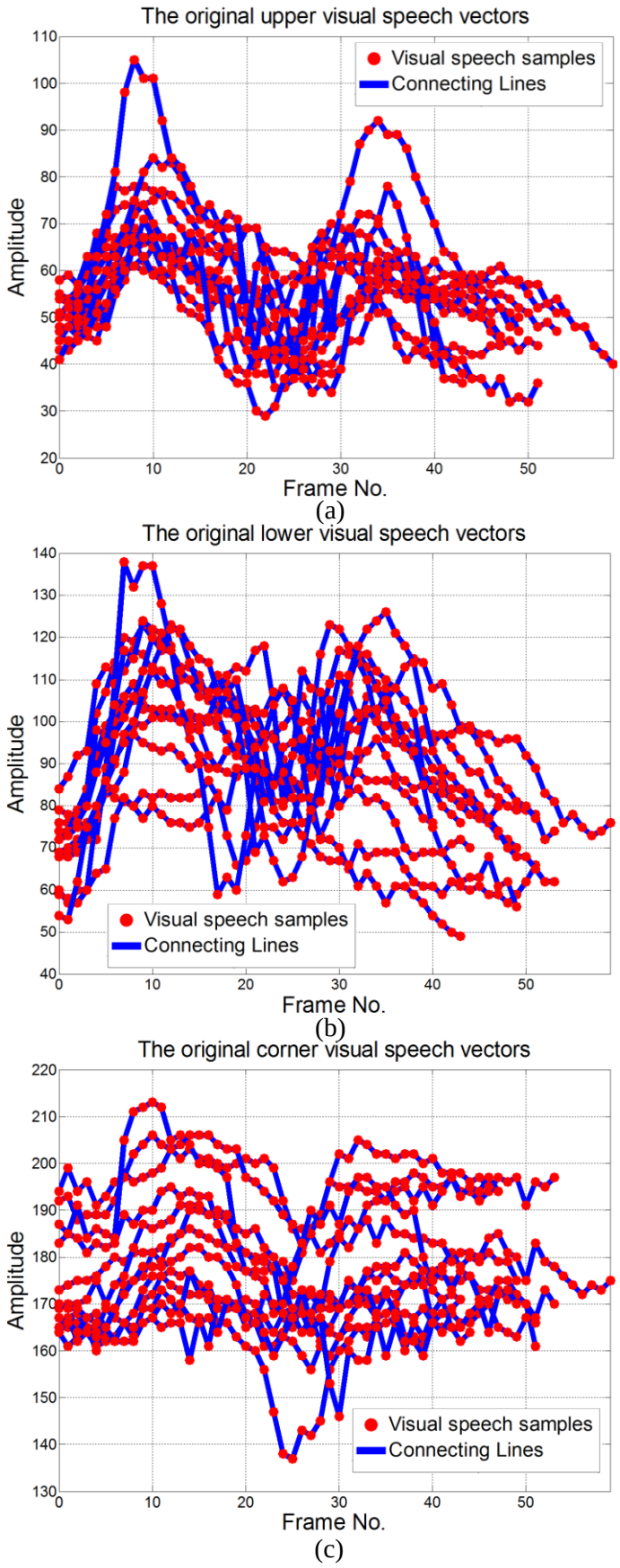


Figure 4-14: The original visual speech sample sets extracted from (a) upper, (b) lower and (c) corner visual features

The scattering plot of samples by allocating upper visual speech samples to x and lower visual speech samples to y axis are shown in Figure 4-15 (a). The lower visual speech samples and corner visual speech samples are shown in Figure 4-15 (b) and Figure 4-15 (c).

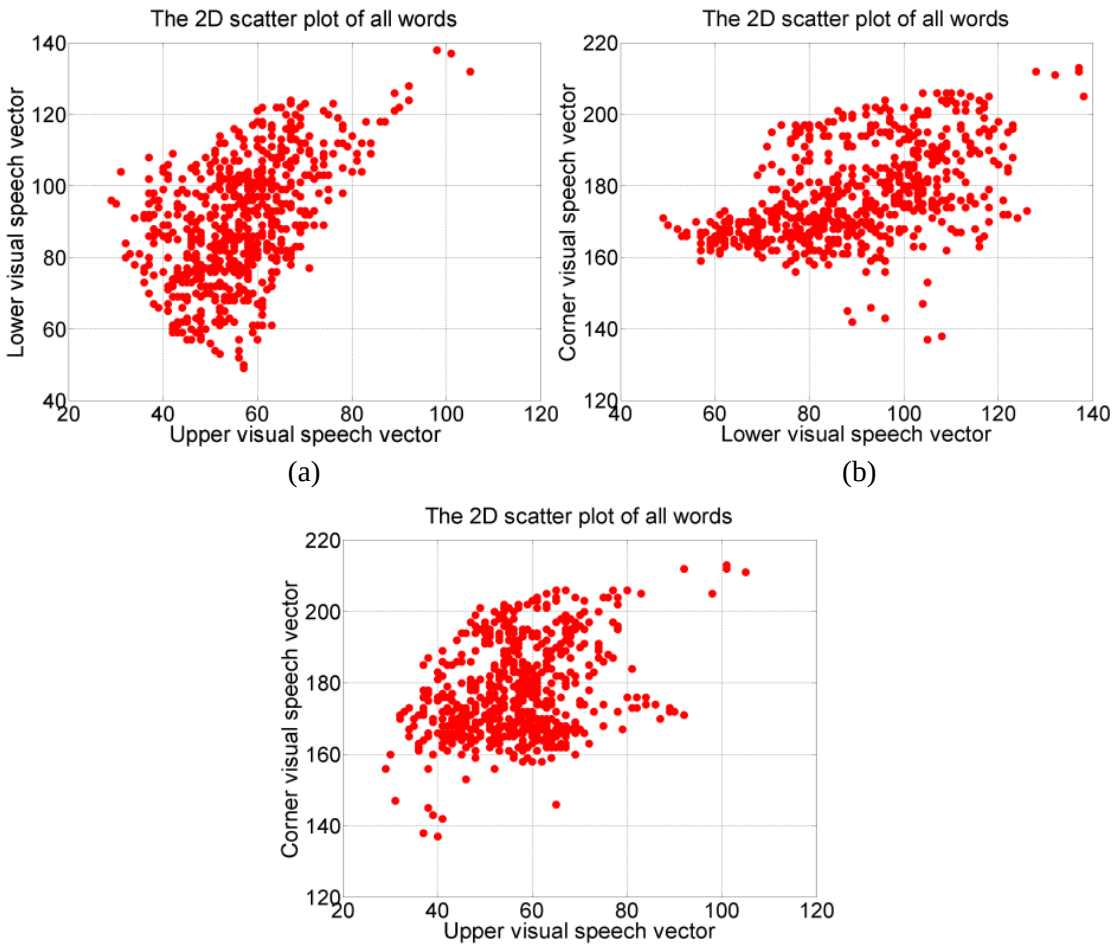


Figure 4-15: The scattering plot of the original visual speech sample sets extracted from upper (a), lower (b) and corner (c) visual features

Obviously, the extracted visual speech samples could not accurately represent the corresponding feature points on actual lip movement. This is due many factors such as speakers head movement, error in marking and extracting the visual speech samples caused by manual extraction. The speakers' head movement in Figure 4-14 can be observed since the starting visual speech sample sets are not equal. The next important issue is the imbalance in the starting and stopping samples. In the next chapter, these issues will be solved by suggesting a novel method.

- **REPRESENTATION OF THE VISUAL SPEECH SAMPLE SETS WITH THE KRONERCKER DELTA FUNCTION**

In this section, the sample sets are represented according the signal processing theory. Regardless to the change in dimension of lips, the samples as they appear in time are arranged in discrete-time domain. Each visual speech sample set consists of F_{W_m} variable numbers of frames.

Using Eq. (2-20), the sample sets ($FP^{u}_{W_m}[i], FP^{l}_{W_m}[i], FP^{c}_{w_m}[i]$) can be modelled by Dirac delta functions. As described before, the shifted versions of Dirac function that are scaled by feature point samples define the DT signals and can be represented as:

$$u^{W_m}[n] = \sum_{k=0}^{F_{W_m}-1} u_i^{W_m}[k]\delta[n-k] \tag{4-20}$$

$$l^{W_m}[n] = \sum_{k=0}^{F_{W_m}-1} l_i^{W_m}[k]\delta[n-k] \tag{4-21}$$

$$c^{W_m}[n] = \sum_{k=0}^{F_{W_m}-1} c_i^{W_m}[k]\delta[n-k] \tag{4-22}$$

where $u_i^{W_m}[k] = FP^{u}_{W_m}[n]$, $c_i^{W_m}[k] = FP^{c}_{W_m}[n]$ and $l_i^{W_m}[k] = FP^{l}_{W_m}[n]$.

4.5. COMPARISON BETWEEN AUTOMATIC AND MANUAL EXTRACTIONS

In this stage, the video file of articulated word ‘Sympathetically’ has been used for comparison. In other words, the visual sample sets of this word are extracted manually and automatically from speaker 1, and the results are compared to justify the usage of manual extraction. The block diagram in Figure 4-16, describes the manual and automatic visual samples extraction processes.

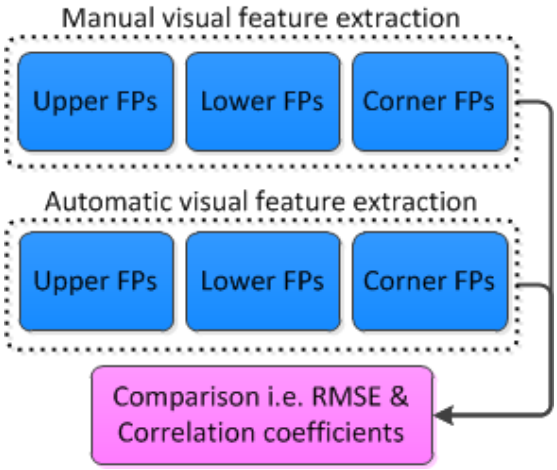


Figure 4-16: Comparing the manual and automatic visual samples extraction

The results of comparing the manual and automatic visual samples extraction from the upper visual feature point is shown in Figure 4-16 (a). The amount of visual similarity is acceptable with the RMSE equal to 1.168 and correlation coefficient 0.887 (Table 4-7).

On the other hand, in Figure 4-16 (b), automatically extracted visual sample sets from the lower feature point neglects important fluctuations between 16^{th} and 25^{th} frames. The RMSE and correlation coefficient between manual and automatic visual sample extractions are equal to 3.704 and 0.577 according to Table 4-7.

For the corner visual feature point, the difference between manual and automatic visual sample extractions is observable; as it shown in Figure 4-16 (c). Furthermore, the RMSE between two methods is highest (5.363) and correlation coefficient is negative (-0.297), which indicate high dissimilarity between the manual and automatic extraction of visual speech samples.

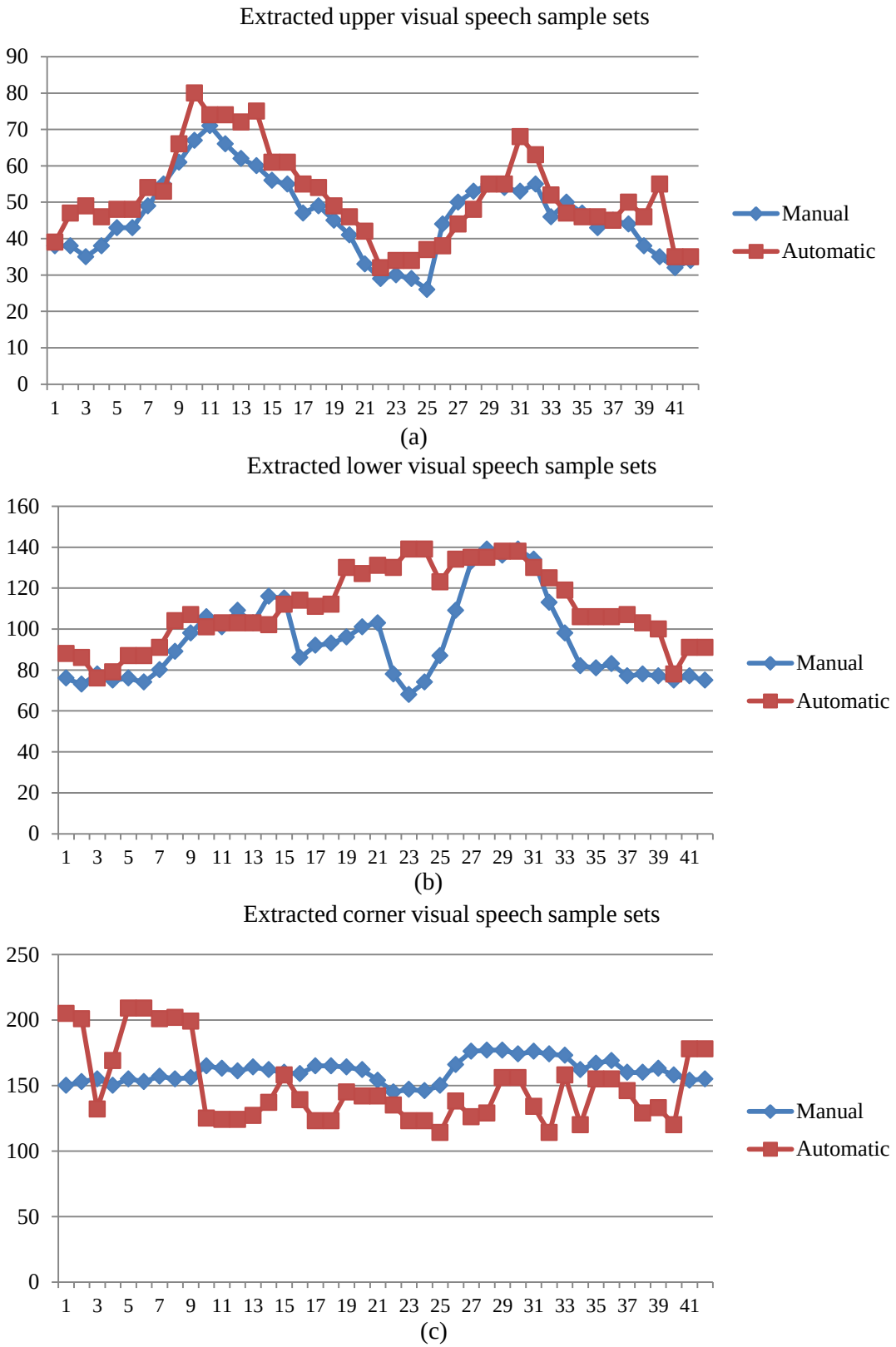


Figure 4-17: The manual and automatic visual sample extraction from (a) upper, (b) lower and (c) corner visual fatures.

Table 4-7: The RMSEs and correlation coefficients between manual and automatic visual sample extraction

	Upper visual sample set	Lower visual sample set	Corner visual sample set
RMSE	1.168	3.704	5.363
Correlation coefficients	0.887	0.577	-0.297

According to the results of comparisons, the manual visual speech sample extraction has been employed in this thesis to increase the accuracy of mathematical expressions of visual speech signals.

4.6. SUMMARY

In this chapter, the methods of parameterizing the lip shape as well as visual data acquisition have been represented. In order to extract the visual information, the rectangular region of interest (ROI) is defined around the lip. Extracting the lip geometry is performed on the ROI manually via three visual features located on the upper and lower outer contour of lip as well as the lip corner. In the next chapter, the visual data acquisition will be performed on the ROI, visual feature points and visual sample sets which are determined in this chapter.

There have been some issues caused by the speakers and the method of lip's movement extractions, which had been addressed. Since non-English speakers articulated the words, it is necessary to obtain the data similar to the English speakers. This method is valid for different accents in English speakers. In order to enhance their manner of articulation, the corresponding audio files have been broadcasted to the speaker. The speakers articulated each word 20 times for further accuracy of visual pronunciations.

The lip motions had been captured during articulating the family of words designed in Section 3.3.4. For increasing the accuracy of articulations, the words audio files have been broadcasted to the speakers. The speakers' video files then have been edited manually to segment each word. In parallel, the time durations of audio files have been extracted and the possible numbers of frames that each word has been approximated. In this stage, the effect of anticipation is considered by adding the constant value to the approximated frame lengths (assuming extra three frames in the

beginning and in the ends). Comparing the manually extracted number of frames and the approximated words frame numbers from audio file, the optimized visual data is detected. The first issue in this process is the possible dissimilarities between the approximated frame lengths from audio files and the manually extracted framesets. Next, there have been repeated visual data sets with equal numbers of frames. The solutions to these issues have been provided. If the approximated number of frames (from audio files) did not match to the manually extracted number of frames, the rounded trimmed mean to the nearest integer of the manually extracted number of frames would be selected as the index of visual speech sample set selection.

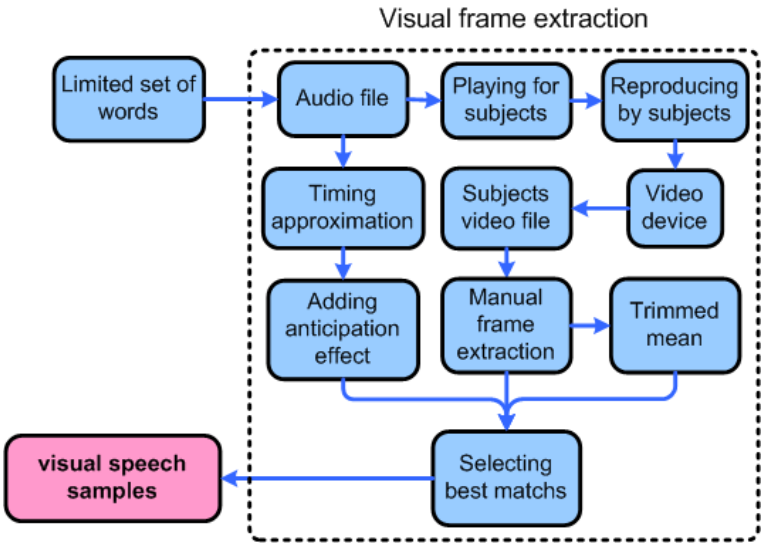


Figure 4-18: The schematic diagram for visual data selection

The amplitudes of those sample sets with equal frame lengths, have been averaged and rounded to the nearest integer (pixel definition) and selected as the visual speech sample set. The suggested methods of visual sample sets selection have been shown in block diagram in Figure 4-18. This method uses the time durations of words audio files and their approximated frames that have been selected manually. However, the automatic visual speech sample extraction could be employed while arbitrary decrease of accuracy depending on the employed algorithms, quality of images and ambient light must be considered. In other words, the human supervision in frame selection and the time duration of words have been joined to choose the best visual representations of the lip's movements. In this work, the sample sets have been represented according to notation in the signal processing theory by delta function.

The final visual word frames have been selected according to the trimmed mean of lengths among all 40 sample sets for each word.

Chapter 5

PROCESSING THE VISUAL SPEECH SAMPLE SETS

In this chapter, the effects of misbalancing the visual speech samples caused by head movement during articulation will be studied and eliminated. Like any acquired data set, it is necessary to apply modification to tackle issues as head movement and face tilting. By considering the source of acquired data (lip region) and the method of extraction, such issues, which affect the ideal observation of the data source, can be addressed.

In this chapter, the issue of lip's displacement during the articulation has been addressed. The speaker's lip as the data source is under the effect of speaker's head position. This is due to unintentional movement of head in all possible direction. Likewise, the method of manual extraction definitely introduces error to the observed data.

More specifically, adding anticipation is purely based on the editor decisions. It means he decides when the lip intended to start and when it stops pronunciation. Furthermore, marking the feature points on each frame could not be done perfectly. The first issue in the visual speech sample sets regards to the level differences (misbalance) between the starting and ending positions of the visual samples. In other words, the lip's position fluctuates due to speakers head movement.

5.1. VISUAL SPEECH DATA MODIFICATION

Modification of a visual speech sample sets is accomplished by adjusting the first and the last feature points on a collinear level with reference to the starting point. During this stage, the data samples need to be translated to the amplitude of starting points. The data samples should be rounded since the pixel values are representing by integers. The next issue is rooted in the speaker's head movement during articulation

of words. It resulted in dissimilarities between staring feature points amplitudes. This issue is adjusted by convergence of the visual speech sample sets to a constant point. After obtaining the adjusted modified visual speech sample sets, the results is concluded by convergence of the upper, lower and corner visual speech sample sets to the mean value of end points amplitudes. The block diagram in Figure 5-1 depicts these processes.

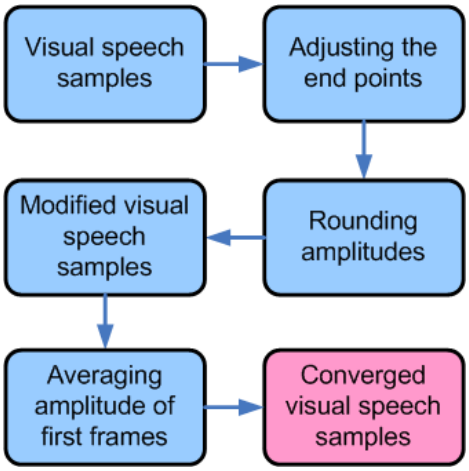
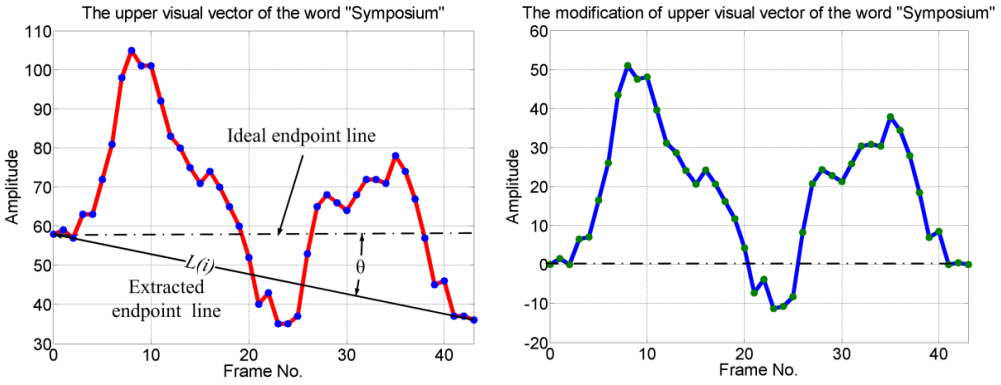


Figure 5-1: Modification and converging of visual speech sample sets

In Figure 5-2 (a), as an example, the upper visual speech sample set extracted from one of the speaker's video files during pronouncing of the word 'symposium', represents the effect of displacement between the starting and ending frames. The degree of end point displacement from an ideal (dashed) line $L(i)$ is denoted by θ. The degree of displacements for upper, lower and corner visual speech sample sets for each word are denoted by $\theta^{u}_{W_m}$, $\theta^{l}_{W_m}$ and $\theta^{c}_{W_m}$, respectively. Therefore, each visual data's degree of displacement θ_{W_m} is:

$$\theta_{W_m} = \begin{bmatrix} \theta^{u}_{W_m} \\ \theta^{l}_{W_m} \\ \theta^{c}_{W_m} \end{bmatrix} \tag{5-1}$$



(a) (b)

Figure 5-2: The extracted upper visual speech sample sets (a) corresponding to word 'symposium' and (b) its adjusted version

The line $L(i)$ shows the amount of deflection by connecting the end points. The deflected visual speech sample set is similar with the effect of adding to signals, one with high frequency and low amplitude, and the other with low frequency and high amplitude. The result of such combination is looking like a fast signal following a path that has been defined by the slower signal. Such effect appears when an added noise (head movement) mixes with a signal. For instance, the upper visual speech sample set for word 'symposium' could be thought as the fast signal and the connecting line between the end points as the slower signal. For example, the extracted upper feature point's visual speech sample set $FP^c_{W_m}(i)$ and the connecting line $L(i)$ are related as:

$$FP^c_{W_m}(i)' = L(i) + FP^c_{W_m}(i) \tag{5-2}$$

where $FP^c_{W_m}(i)'$ is the desirable visual speech samples after modification and $i = 0,1,2,\dots,(F_{W_m} - 1)$. The line $L(i)$ is derived by connecting the first and last sample. Therefore, simply by subtracting the visual speech sample set from the connecting line, the original signal is obtained. The adjusted visual sample set is depicted in Figure 5-2 (b). The amplitude of the $FP^c_{W_m}(i)'$ afterward could brought up to the amplitude of $FP^c_{W_m}(i)'$ by adding a constant value related to trimmed mean of all first sample. The data samples are rounded to their nearest integer and are provided to the next stage of visual speech sample set modification by biasing their amplitudes. The amplitude biasing simply has been performed by adding a constant value to all samples. The biasing constant is defined by averaging the first frames in each data set.

The biasing operations are converged the upper, lower and corner visual speech sample sets to the averaged values of the starting amplitudes. For example, the degree of displacement for the word 'symposium' as defined in Eq. (5-1) is:

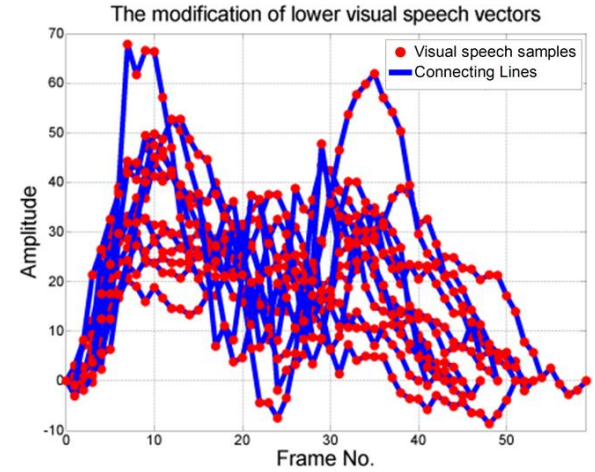
$$\theta_{W_m} = \begin{bmatrix} 26.565^{\circ} \\ -9.039^{\circ} \\ 15.255^{\circ} \end{bmatrix} \tag{5-3}$$

The result in Eq. (5-3) exhibits the displacement of visual features which can also indicate the perfection of articulation. The upper visual feature (upper lip) in the end of articulating the word has an upward displacement, while the lower visual feature (lower lip) has downward displacement. Combination of these two displacements is indicating as if the lip remains slightly open at the end of articulation. The corner visual feature is located on the left corner of speaker lip that has a positive degree of displacement. It also indicates a slight movement of the speaker's face to left.

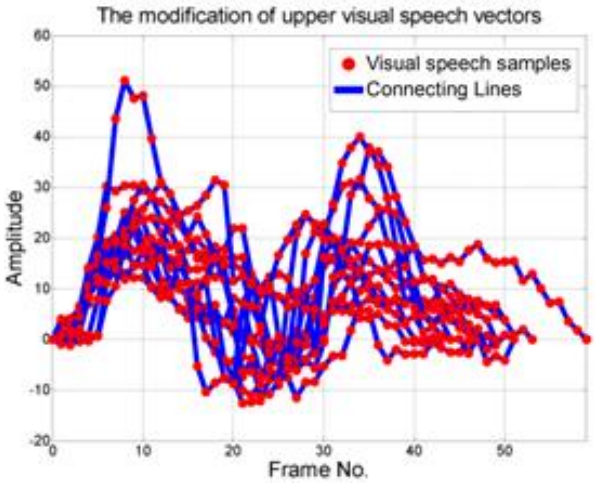
Table 5-1: The degree of displacement in visual data sets

		Upper visual speech sample set misbalance $\theta^{u}_{W_m}$	Lower visual speech sample set misbalance $\theta^{u}_{W_m}$	Corner visual speech sample set misbalance $\theta^{c}_{W_m}$
Visual data	Simplest	3.179830145	2.121096395	-5.290081229
	Simpler	1.145762838	2.290610043	-2.290610043
	Simply	11.76828891	-1.193489405	-8.297144951
	Simple	2.385944049	4.763641672	0
	Sympathetically	10.3888578	-6.654425065	-5.710593137
	Symposium	26.56505118	-9.039482798	15.25511872
	Symptom	8.841814585	-3.814074853	5.079607866
	Symbolism	4.398705359	-6.581944633	-2.202598135
	Symbolize	8.746162271	-1.101706102	0
	Symbolic	3.179830145	6.34019174	-5.290081229
	Symbols	-3.433630362	3.433630362	-5.710593137
	Simmer	1.397181033	-5.572197826	5.572197826
	Cement	-7.765165998	6.483073713	-5.194428913

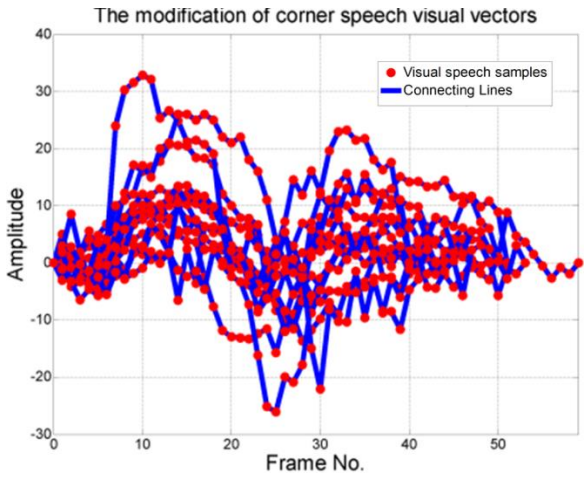
The degree of misbalancing is shown in Table 5-1 where the maximum change in degrees again appears in the word 'symposium'. The upper visual feature is gone upward with almost 26.5° and the lower visual feature is gone down word with almost 9° while the corner visual speech feature is gone to the right since it is located on the left corner of the speakers' lips when the articulations of this word are finished.



(a)



(b)



(c)

Figure 5-3: The (a) upper (b) lower and (c) corner visual speech sample sets after adjusting the end points

The visual speech samples misbalancing is eliminated Figure 5-3, after the modification. The starting samples in all sample sets have same values. In addition,

the first and the last samples for each word have same values. In other words, they are co-linear on the x axis.

Table 5-2: Calculation of biasing values for the visual sample sets

	Average	Rounding
Upper visual speech first sample	49.1538	49
Lower visual speech first sample	69.6154	70
Corner visual speech first sample	173.6923	174

In Figure 5-4 (a) and (b) the upper $u^{W_m}[n]$, lower $l^{W_m}[n]$ and corner $c^{W_m}[n]$, $n = \{0, 1, 2, \dots, F_{W_m} - 1\}$ visual speech sample sets are biased according to the values in Table 5-2 in frame and time domains. These sample sets are sharing starting points. It means the visual features on lip do not appear in different locations and the ROI is fixed.

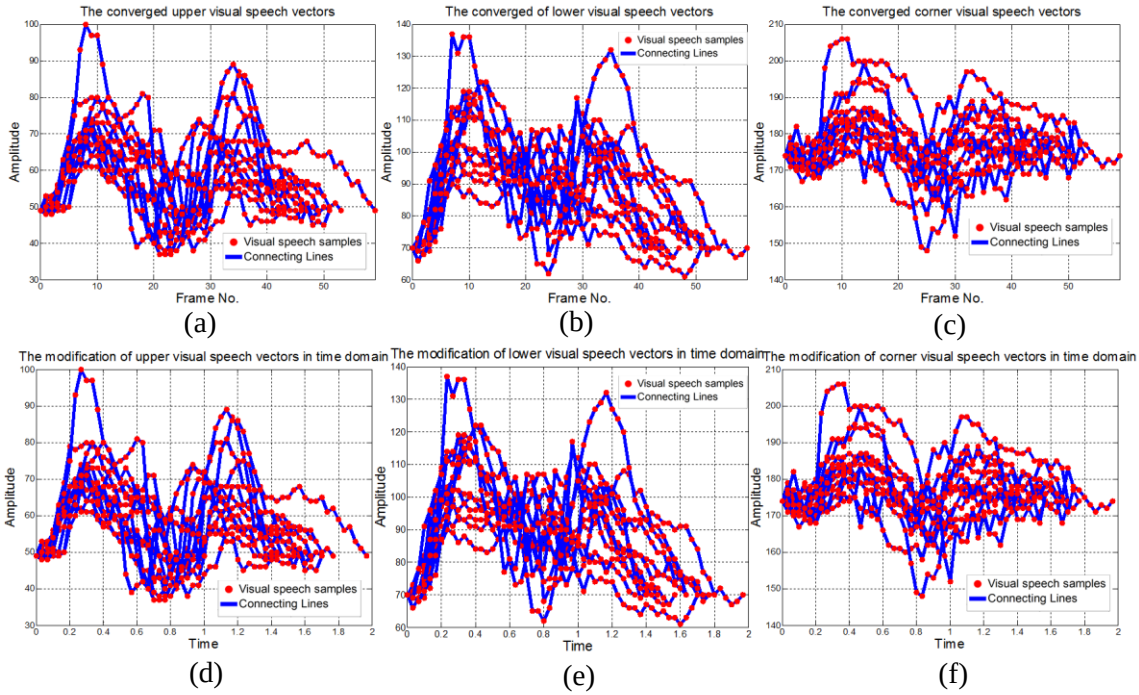


Figure 5-4: The (a) upper, (b) lower, and (c) corner visual speech sample sets in frame domain in time domain (d), (e) and (f) after converging to the corresponding first samples represented in Table 5-2

In Table 5-3, the boundaries of visual speech sample sets in terms of minimum and maximum amplitudes are represented.

$$Min_{W_m}^{u} < FP_{W_m}^{u}(i) < Max_{W_m}^{u}$$

$$Min_{W_m}^{l} < FP_{W_m}^{l}(i) < Max_{W_m}^{l} \quad (5\text{-}4)$$

$$Min_{W_m}^{c} < FP_{W_m}^{c}(i) < Max_{W_m}^{c}$$

Table 5-3: The maximum and minimum amplitudes of the visual speech samples

		$Min_{W_{m_q}}^{u}$	$Max_{W_{m_q}}^{u}$	$Min_{W_{m_q}}^{l}$	$Max_{W_{m_q}}^{l}$	$Min_{W_{m_q}}^{c}$	$Max_{W_{m_q}}^{c}$
Words	Simplest	38	69	68	117	160	184
	Simpler	43	63	70	103	169	186
	Simply	37	70	62	113	158	180
	Simple	40	80	67	105	162	186
	Sympathetically	48	77	67	122	167	194
	Symposium	38	100	70	137	165	206
	Symptom	43	75	69	122	170	200
	Symbolism	38	81	70	119	148	189
	Symbolize	41	73	61	93	167	181
	Symbolic	45	89	70	132	162	184
	Symbols	42	69	68	117	152	184
	Simmer	49	81	70	106	169	187
	Cement	39	64	69	114	162	176

In Table 5-4, the words with overall maximum and minimum visual samples from Table 5-5 are represented.

Table 5-4: The overall maximum and minimum amplitudes of the visual speech sample

	Values	Words
$Min_{W_m}^{u}$	37	Simply
$Max_{W_m}^{u}$	100	Symposium
$Min_{W_m}^{l}$	61	Symbolize
$Max_{W_m}^{l}$	137	Symposium
$Min_{W_m}^{c}$	148	Symbolism
$Max_{W_m}^{c}$	206	Symposium

- **ERROR AFTER ADJUSTING VISUAL SPEECH SAMPLE SETS END POINTS**

Adjusting the visual speech samples end points causes to reducing the speakers head movements and imperfection of lip dimension in the rest positions. The differences between the starting and stopping samples for the visual data have been demonstrated in Figure 5-5. In this figure the visual speech sample sets $u^{W_m}[n]$, $l^{W_m}[n]$ and $c^{W_m}[n]$

in the word ‘symposium’ have the maximum difference between the starting and stopping samples amplitudes.

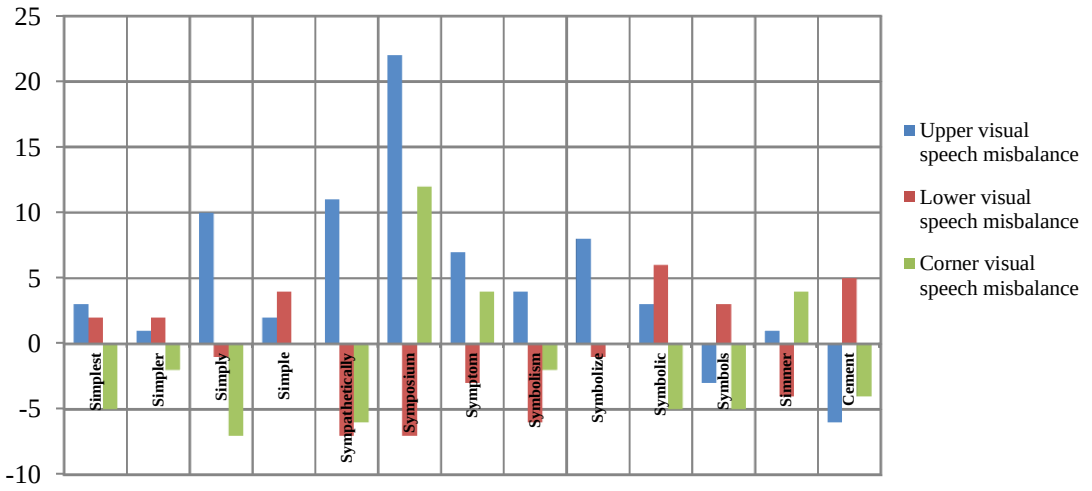


Figure 5-5: Amplitude difference between the last samples

It is also possible to interpret the visual data sample sets misbalancing by calculating the degree of error from the horizontal lines that are defined by the first sample in each sample set. Converging values of the visual data sample sets is represented in Table 5-2. In Table 5-5, the absolute difference values between the first samples in the visual speech sample sets and the converging values are represented as adding constant biasing values.

Table 5-5: The values of biasing in Table 5-2 on the visual data sample sets

		Absolute values of differences between biasing and the first samples in each visual data				
		Upper visual speech	Lower visual speech	Corner visual speech	Mean	Standard deviation
Visual data	Simplest	1	6	18	8.333	8.737
	Simpler	1	2	1	1.333	0.577
	Simply	6	9	13	9.333	3.512
	Simple	1	14	20	11.667	9.713
	Sympathetically	2	1	5	2.667	2.082
	Symposium	9	1	9	6.333	4.619
	Symptom	4	1	10	5.000	4.583
	Symbolism	6	4	9	6.333	2.517
	Symbolize	1	10	10	7.000	5.196
	Symbolic	5	2	9	5.333	3.512
	Symbols	2	11	9	7.333	4.726
	Simmer	8	2	4	4.667	3.055
	Cement	2	16	7	8.333	7.095

The word ‘simple’ has the maximum standard deviation. It indicates the displacement of lip position is at its maximum in comparison to the other visual data. The suggested method corrects the displacement of rest points before and after speaker’s articulation. Therefore, the effect of unintentional speakers head movements is eliminated. This will also reduce the human error in extraction of samples.

For further analysis of data, it is necessary to have ability on modification of the range and domain of the sample set. In this work, the suggested process sets a fixed framework for observation of the visual speech sample sets in order to gain more efficiency standardizing the visual words. The two dimensions of the visual speech sample sets are the subject of transformation to a fixed domain or range. In this work, the operation of transforming range or domains of visual speech sample sets, polynomials, or signals is referring as normalization. This would be beneficial in observation of trend of data.

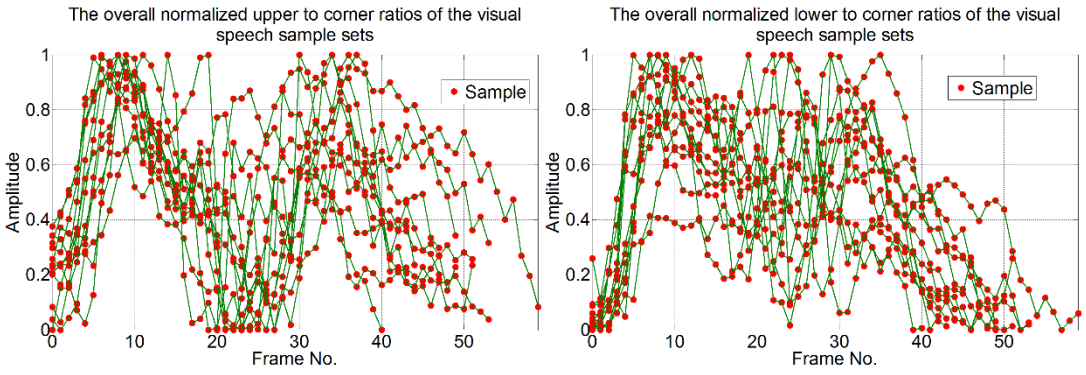
5.2. NORMALIZATION

Normalizing a set of data can be thought as fitting the dimensions of the sample set or signal to a new reference dimensions. The dimensions of the data are amplitudes (ranges) and the intervals (domains). Often the reference dimension is chosen between

zero and one. By scaling normalized signal's range or domain to the unit interval it is possible to fit the signals elements (samples) to a desirable range or domain. This concept is employed for data processing in order to classify the extracted visual speech sample sets. More specifically, it is used in processing the relating of visual speech signals to each other in each word, barcode generation as well as coding the visual speech signals signatures.

5.2.1. AMPLITUDE NORMALIZATION

In the amplitude normalization, the range of data or signal based on the statistical properties of data could be transferred to a fixed and limited range. The range of signal is defined between the minimum and maximum of samples amplitudes. Comparing each sample with the range of all samples in a set is the index of normalization. The comparison index is the ratio of each sample's amplitude to the range of sample set. The signal in this stage has been normalized but there is an offset equal to the minimum amplitude in the sample set that prevents it to be located between zero and one. Eliminating the effect of such offset simply can be done by subtracting the minimum amplitude from all amplitudes.



(a) (b)

Figure 5-6: The normalized (a) upper to corner ratios and (b) lower to corner ratios of the visual speech sample sets for all 13 words

In order to relate the visual speech signal (Section 6.2), the visual speech sample sets are related for purpose of the visual word unification. The relation is based on the ratio of the visual speech sample set. In Figure 5-6, the upper and lower visual speech sample sets are divided by the corner visual speech sample sets and then are normalized.

5.2.2. DOMAIN NORMALIZATION

Normalizing the sample set's domain is also similar to the amplitude normalization. It is desirable to transform a sample set's domain to anther interval mostly between zero and one. If the sample set's domain is defined on the x axis discretely $[0: N-1]$, the span x_i will be define as $x_i = i, i = 0, 1, 2, \dots, (N-1)$, where $N \in \mathbb{N}$ is the number of samples. It means an interval between zero and $N-1$ is uniformly spanned by one has been defined i.e. $(x_{i+1} - x_i = 1)$. It is possible to provide more flexibility on boundary limits and controlling the expansion and reduction of span distances by normalizing the interval by multiplying $x_i = i$ with a factor $\frac{1}{(N-1)}$. Therefore, by dividing the normalized span by a factor β, the number of spans can be changed. If $\beta \to \infty$, then a continuous interval $x_\beta = [0, N-1]$ is:

$$x_\beta = \{{x_i}^\beta\} \qquad {x_i}^\beta = \frac{i}{\beta(N-1)} \qquad i \in \mathbb{N} \tag{5-5}$$

The expression (5-5) associates the ability of increasing data set domain to a particular domain $[0: (N-1)]$ that is spanned by the factor $\frac{1}{\beta}$. In other words, the domain of sample sets would be increased by inserting extra equidistance spans. The numbers of inserted spans, which are considered as the fundamental operation in the interpolation, are $\beta - 1$. In this stage, the statistical characteristics of visual data before and after modification are desirable to enhance the knowledge about their source. Afterward, the lip as the source of extracted data during utterance of chosen words can be categorized in a statistical model. The geometry of the data source is hidden inside the density functions estimated from visual speech sample sets. Although there is no information provided about the sequence of samples but it could estimate the number of samples where lips geometry represent in the most of visual data frames. The lip's statistical model is also used as a pattern for digital representation of visual speech signals.

5.3. RELATIONS OF VISUAL SAMPLES AND ARTICULATION MANNER

In this stage, it is possible to study the dynamic of lip for articulating the words. Analysing the visual speech samples is possible from two perspectives:

- Analysing the visual speech sample sets trends
- Analysing the visual speech sample sets trends via phonemic structures defined according to the transcribe words

In the first approach, the results of analysing is described by a single trend for each upper, lower and corner visual speech sample sets. It means the upper sample sets can be described by a single trend. The lower and corner visual speech sample set also have a single trend. On the other hand, analysing the visual speech sample sets according to the phonemic structure of words, suggest different trends for groups of words categorized by appearing phonemes sequentially. Both methods using a segmentation approach based on the overall trend of sample sets. Both aspects are studied and the results in which viseme are grouped according to the place and manner of articulation are represented and interpreted using a phoneme-viseme table.

In order to have more comprehensive and clear observation of visual speech sample set behaviour, averaging of sample sets will be considered and processed further by normalization. This process is shown in **Figure 5-7**. The averaging of all samples could not be a perfect candidate for representing the visual speech sample sets trend since the length of each visual speech sample sets do not have same length. Therefore, the maximum number of samples included in this process is limited to 41 samples since it is belong to shortest visual data in articulating the word 'simmer'.

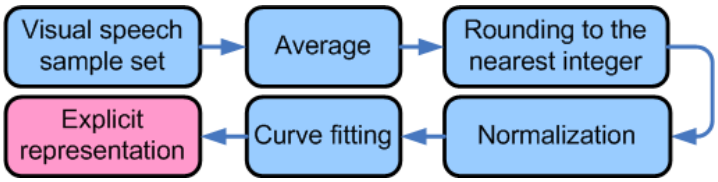


Figure 5-7: The visual speech sample sets analysis scheme for approximating their trend

The result of suggested approximation is shown in Figure 5-8. The result of approximation shows dominance trend of data by averaging. For example, in Figure 5-8 (a) the sample amplitudes appeared bellow the starting of upper visual feature that means the upper lip is a closure trend. However, this effect is hidden by other amplitudes in averaging process in Figure 5-8 (b) and (c).

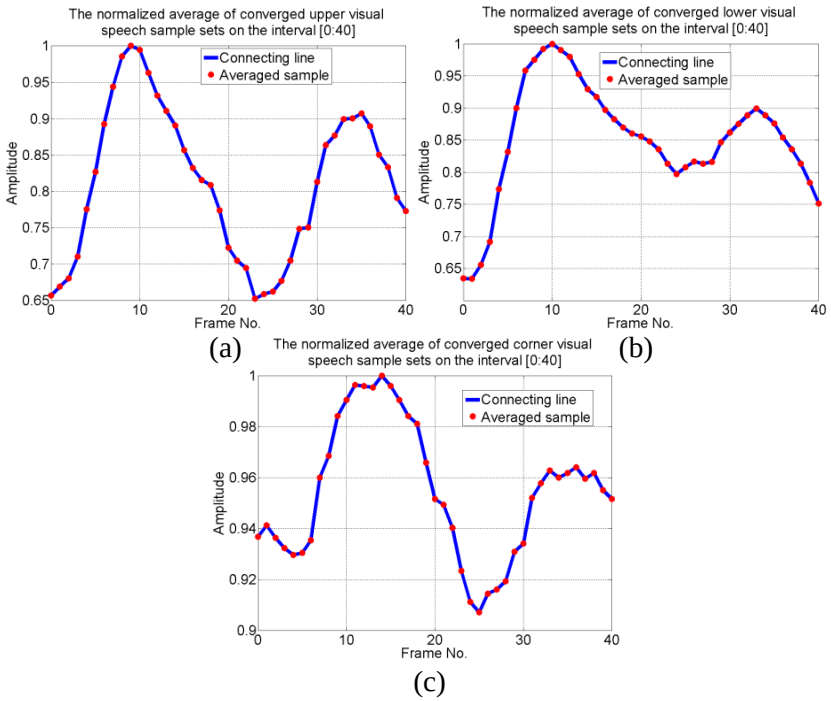


Figure 5-8: The normalized averaged of (a) upper, (b) lower and (c) corner visual speech samples on the interval $\boldsymbol{F_w} = [\mathbf{0}: \mathbf{40}]$

The first observation from Figure 5-8 suggests a coherence incremental pattern. It is shared with upper and lower visual speech sample sets for $F_W \in [0:10]$ or from 0 to 379 milliseconds. On the other hand, the corner visual speech sample sets show more steady incremental change. The second segment starts in 12th frame until the 25st frame where almost all visual sample sets obtained by upper, lower and corner visual features reach to their minimum. From the 24st frame until the 34th frame, the trends of visual speech sample sets in all three cases again increasing. Afterward, from 35nd frame until the ending frame in each sample set, the trends are decreasing. The visual speech samples are sharing similarities in a portion of beginning frames. This is because the words phonemic structure has same sequence of three phonemes (/s/+/ih/+/m/+…). Therefore, to avoid this issue, based on approximation of numbers of frames, which are covering first three phonemes, the averaging performed on first 21 samples. The segment approximation differs regarding to reliability of framework which is used to interpret the lip movement.

In this study, the answer is suggested by selecting a phoneme-viseme table. According to the phoneme-viseme table used by (Neti, et al., 2000) as it represented in Table 5-6 it is possible to approximate the lip dimensions. In Figure 2-1, the place of articulation is also represented.

Table 5-6: The phoneme-viseme table (Potamianos, et al., 2004)

Viseme	Phoneme
Silence	/sil/, /sp/
Lip-rounding based vowels	/ao/, /ah/, /aa/, /er/, /oy/, /aw/, /hh/, /uw/, /uh/, /ow/, /ae/, /eh/, /ey/, /ay/, /ih/, /iy/, /ax/
Alveolar-semivowels	/l/, /el/, /r/, /y/
Alveolar-fricatives	/s/, /z/
Alveolar	/t/, /d/, /n/, /en/
Palato-alveolar	/sh/, /zh/, /ch/, /jh/
Bilabial	/p/, /b/, /m/
Dental	/th/, /dh/
Labio-dental	/f/, /v/
Velar	/ng/, /k/, /g/, /w/

The phoneme /s/ is categorized as 'alveolar-fricatives' in the visual domain. The tongue blade against alveolar rigid (place of articulation) pronounced it. Obviously, the lips dimension increases to reach a position of articulating the next phoneme. The next phoneme appearing in all words in the second position of their sequential phonemic structure is /ih/. The corresponding visual domain representation of this phoneme is 'lip-rounding based vowels'. The lip dimension has also an incremental trend. Segmenting these two phonemes separately is not essential since as mentioned, the Co-articulation effect in the visual domain is included in this work. Therefore, the combination of them as /s/+/ih/ in the visual domain can be expressed as a constant increment of lip dimension. The third phoneme /m/ is grouped as bilabial viseme. As mentioned in Section 2.2, the bilabial viseme is visual correspondence of bilabial phonemes. The places of articulating phoneme /m/ is upper and lower lip when they converge toward each other. This means in Figure 5-8, visual phoneme /m/ appears in the first maximum and ends in the first minimum amplitudes between $F_W \in [9:24]$. This means the phoneme visually appears around 10^{th} frame and approximately stops in 25^{th} frame with decreasing trend in lips dimensions. Immediately after articulating phoneme /m/ in frame 26, amplitudes of the upper, lower, and corner visual speech sample sets increases and lasts until 34^{th} frame where they reach the second maximum. It means during these frame lip is opening. From the second maximum amplitude until 41^{st} frame, the overall shape of lip is tending to closure. By using sum of sinusoidal function in MATLAB curve fitting toolbox, the corresponding

polynomial outcomes of averaged normalized upper, lower, and corner visual speech sample sets are depicted in Figure 5-9.

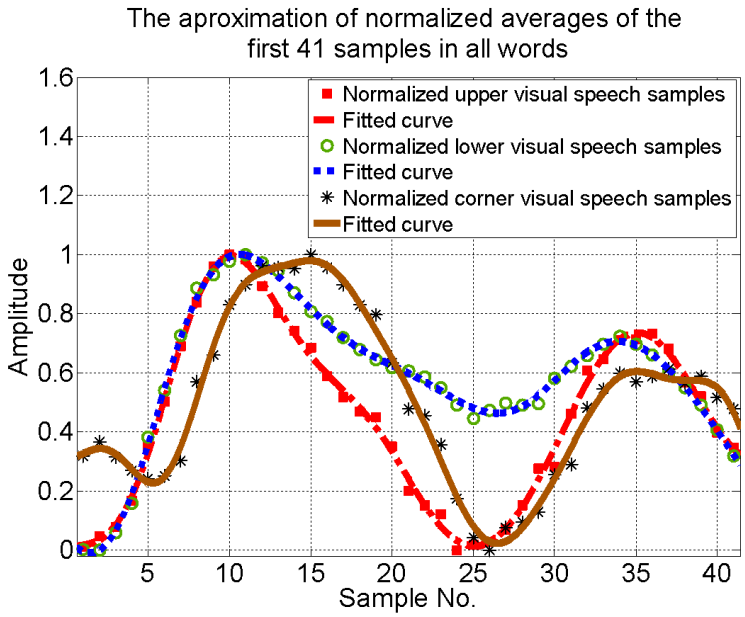


Figure 5-9: The approximated polynomials of the upper, lower and corner visual speech sample sets on the interval $\boldsymbol{F_W} \in [\mathbf{0}: \mathbf{40}]$

$$S^U(x) = 0.864sin(0.08949x - 0.5412) + 0.6862sin(0.1638x + 0.6674) + 0.1726sin(0.338x - 3.124) + 0.118sin(0.5048x - 3.193) + 0.04791sin(0.6275x + 2.666) + 0.007047sin(0.789x + 0.3688) \quad (5\text{-}6)$$

$$S^L(x) = 1.1sin(0.08521x - 0.3862) + 0.5476sin(0.1547x + 0.9073) + 0.1469sin(0.3702x - 3.728) + 0.2357sin(0.4967x - 3.617) + 0.1586sin(0.5562x + 4.145) + 0.03408sin(0.6578x + 4.886) \quad (5\text{-}7)$$

$$S^C(x) = 0.7623sin(0.06773x + 0.3193) + 0.3676sin(0.2848x + 3.532) + 0.2301sin(0.1359x + 1.511) + 0.05235sin(0.6149x + 1.102) + 0.1581sin(0.9068x - 2.425) + 0.1587sin(0.9251x + 0.2361) \quad (5\text{-}8)$$

where variable x allocated to samples.

Table 5-7: The goodness of fitting for the approximating the normalized averaged of the first 41 samples in all sample set

		$S^U(x)$	$S^L(x)$	$S^C(x)$
The sum of sine functions approximation Goodness of fit	SSE	0.01873	0.008633	0.03349
	R-square	0.9948	0.9965	0.9898
	Adjusted R-square	0.991	0.9939	0.9822
	RMSE	0.02854	0.01937	0.03816

By changing the perspective of analysis, the phonemic structure of words is used for describing the visual speech sample sets trends. Referring to Figure 3-5, the three levels of the decision tree of words' phonemic structure has been studied. The result proves the similarities of visual speech sample sets trends for the beginning portion of sample sets.

Analysing the dynamism of lip during articulation from this point divides to possible branches that are occurring in each level according to Figure 3-5. According to these branches, it is desirable to observe similarities in four groups of words that are the one who share /p/, /b/, /ax/ and /eh/ in the fourth phonemic positions. In Figure 5-10, these words are categorized. In the 4^{th} level seven words sharing phoneme /p/, four of them sharing phoneme /b/, one has phoneme /ax/ and one has phoneme /eh/. Regarding to these categories, the word visual speech sample sets are studied.

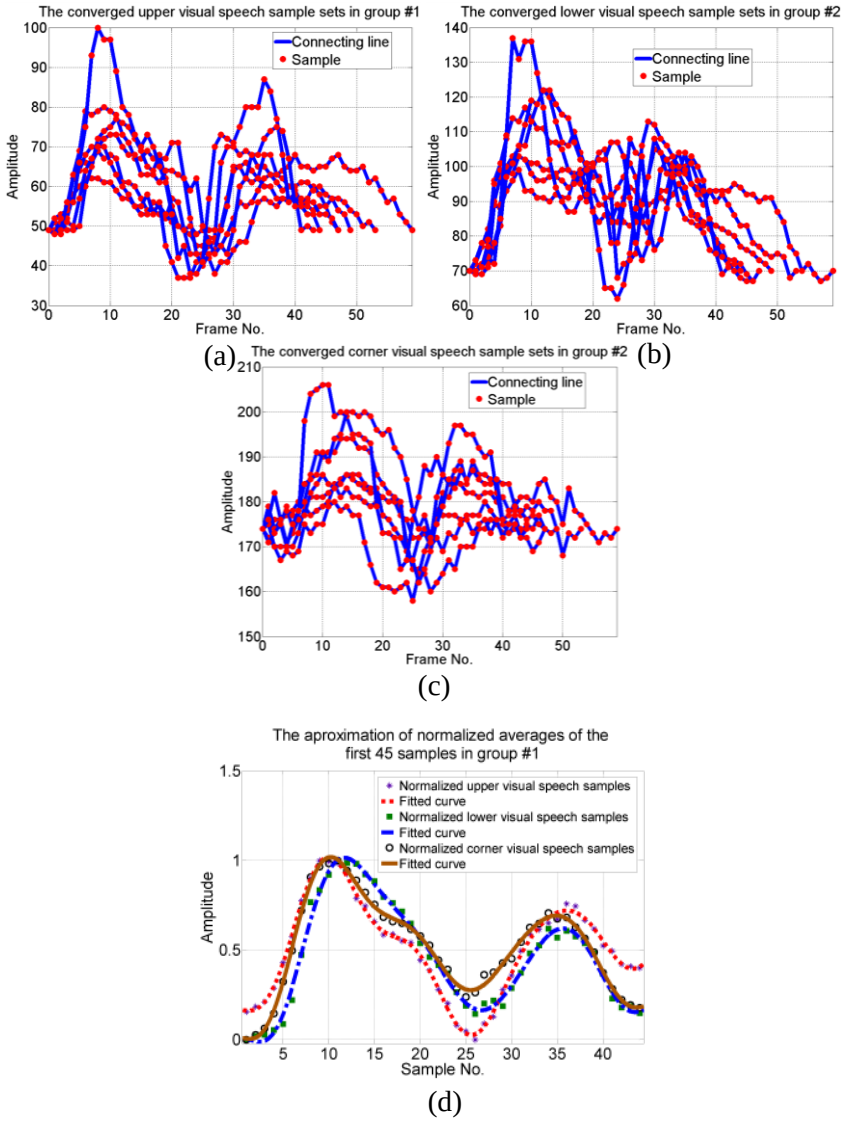


Figure 5-10: The (a) upper, (b) lower and (c) corner visual speech sample sets in the first group of the selected words and the (d) approximated polynomials

The first group of words consists of 'simplest', 'simpler', 'simply', 'simple', 'sympathetically', 'symposium', and 'symptom'. In this group, the phoneme sequence /s/+/ih/+/m/+/p/ is in common. As it described previously and represented by the approximated normalized average polynomials in Figure 5-10 (d), it can be assumed that the first 26 samples of the visual features are representing the phoneme combination /s/+/ih/+/m/. Continuing from the 27nd frame, the upper and corner visual sample sets show an increasing trend more or less until the 35th frame where the second local maximum occurs. For the lower visual sample sets, such trend is also clear where there is an increasing trend around 27th until 35th frame. According to Table 5-6, the phoneme /p/ is categorized as a bilabial viseme. It means the place of articulation is upper and lower lip where phoneme /p/ starts from minimizing the

distance between them and forcing air through them and ending in slightly opened lip. Comparing the visual speech sample sets in Figure 5-10 (a) to (c), the approximated polynomials in Figure 5-10 (d), shows the last section of samples starting in 36th frame has a decreasing trend. The approximated polynomials from the normalized average of the upper, lower and corner visual speech sample sets in this group are:

$$S_1{}^U(x) = 1.016sin(0.07602x - 0.4071) + 0.7406sin(0.1313x + 1.232) + 0.2068sin(0.3059x - 2.519) + 0.06208sin(0.6243x + 1.996) + 0.111sin(0.4812x - 2.562) \quad (5\text{-}9)$$

$$S_1{}^L(x) = 0.8334sin(0.0864x - 0.4364) + 0.1496sin(0.3709x - 4.24) + 0.5265sin(0.177x + 0.1034) + 1.086sin(0.5505x + 4.205) + 1.164sin(0.5434x - 5.014) \quad (5\text{-}10)$$

$$S_1{}^C(x) = 1.005sin(0.08985x - 0.4848) + 0.1356sin(0.3844x - 4.139) + 0.6086sin(0.1681x + 0.5951) + 0.1743sin(0.4991x - 3.4) + 0.08974sin(0.5899x + 3.354) \quad (5\text{-}11)$$

Table 5-8: The goodness of fitting for the approximating the normalized averaged of the first 44 samples in the first group

		$S_1{}^U(x)$	$S_1{}^L(x)$	$S_1{}^C(x)$
The sum of sine functions approximation Goodness of fit	SSE	0.01621	0.04139	0.01562
	R-square	0.9948	0.9891	0.9949
	Adjusted R-square	0.9923	0.9839	0.9925
	RMSE	0.02364	0.03778	0.02321

The second group of words is combined of four visual sample sets with the common phoneme combination /s/+/ih/+/m/+/b/. The members of this group are ‘symbolism’, ‘symbolize’, ‘symbolic’ and ‘symbols’. The visual speech sample sets of them are depicted in Figure 5-11 (a), (b) and (c). The approximated polynomials of the normalized average of the first 50 samples are illustrated in Figure 5-11 (d). The visual data in this group act differently from the first group. Although there are two regions for local maximum amplitudes appeared, the first one locate between 10th and 11th frames has included lower amplitudes than the second local maximum region for the upper and lower visual speech sample sets. This behaviour indicates the phoneme, which its articulation end in the second local maximum, is produced by wider dimension of lips in compare with the phoneme combination /s/+/ih/. Similar to the second group of visual data, the first 24 samples are representing the frames that lip is articulating /s/+/ih/+/m/. On the other hand, the corner visual speech sample sets on this interval show smoother trends. From 25th sample until the 39th sample the second

local maximum is occurred. On this interval, the phoneme /b/ is articulated. According to Table 5-6, the phoneme /b/ is categorized in same group of phoneme /p/ as bilabial visual phoneme but represents a wider upper and lower lip dimension. The other noticeable configuration behaviour, which is appeared in both upper and lower visual speech sample sets on this interval, is a similar decreasing trends with different slops. The word 'symbolize' has the steeper decreasing trend followed by the words 'symbolic', 'symbolism' and 'symbols'. The length of these visual data is approximately equal. The corner visual speech sample sets after reaching to first local minimum amplitude return to the rest position.

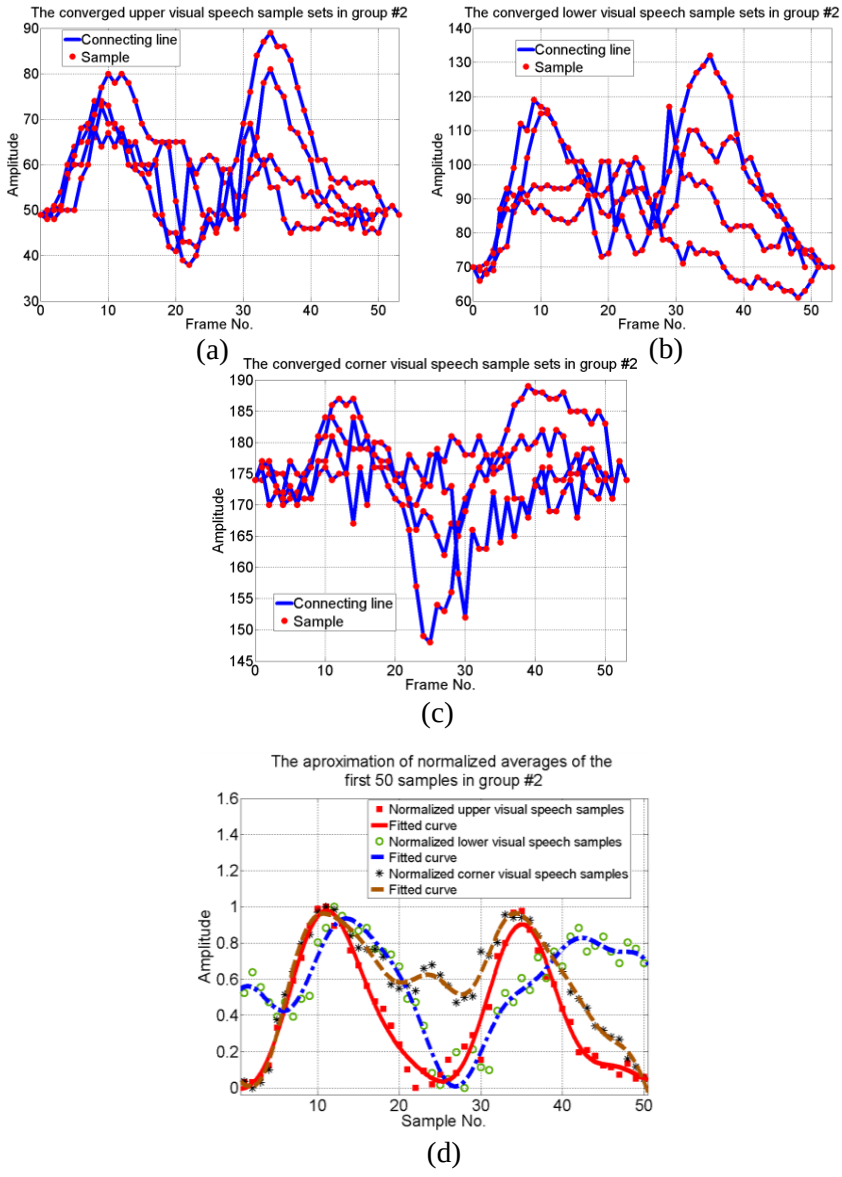


Figure 5-11: The (a) upper, (b) lower and (c) corner visual speech sample sets in the second group of the selected words and (d) the approximated polynomials

The polynomials, which are shown in Figure 5-11 (d), are approximated by the normalized ratio of averaged visual speech sample sets. These polynomials are represented by sum of sinusoidal function:

$$S_2{}^U(x) = 0.5722sin(0.05084x + 0.2962) + 0.4081sin(0.2653x - 1.529) + 0.1784sin(0.1214x + 1.406) + 0.8285sin(0.4826x + 3.081) + 0.7516sin(0.4779x + 0.01326) \quad (5\text{-}12)$$

$$S_2{}^L(x) = 1.413sin(0.09787x + -0.8846) + 1.793sin(0.1647x + 0.5346) + 1.89sin(0.2449x - 4.485) + 1.39sin(0.2732x + 4.152) + 0.05933sin(0.6215x + 0.614) \quad (5\text{-}13)$$

$$S_2{}^C(x) = 0.7826sin(0.05189x + 0.3266) + 0.06753sin(0.3883x - 3.541) + 0.2552sin(0.2419x - 1.158) + 0.1207sin(0.4836x + 3.301) + 0.04147sin(0.7435x + 2.275) \quad (5\text{-}14)$$

Table 5-9: The goodness of fitting for the approximating the normalized averaged of the first 44 samples in group 2

		$S_2{}^U(x)$	$S_2{}^L(x)$	$S_2{}^C(x)$
The sum of sine functions approximation Goodness of fit	SSE	0.1622	0.244	0.0733
	R-square	0.9691	0.9291	0.9815
	Adjusted R-square	0.9567	0.9008	0.9741
	RMSE	0.06807	0.0835	0.04576

The third group consists of one word ‘simmer’. The visual speech sample sets with the phoneme combination /s/+/ih/+/m/+/ax/ is depicted in Figure 5-12.

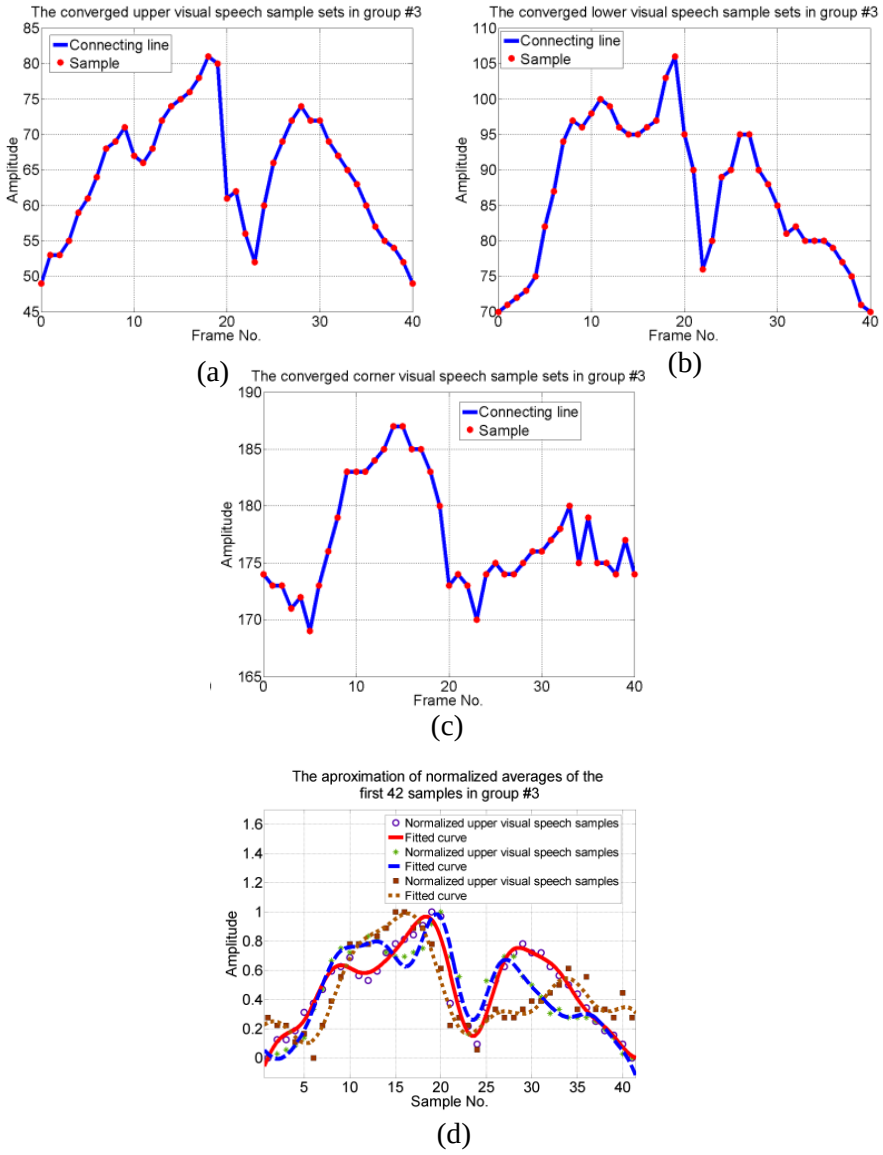


Figure 5-12: The (a) upper, (b) lower and (c) corner visual speech sample sets in the third group of the selected words and (d) the approximated polynomials

All the visual sample sets, which is allocated to articulating the phoneme sequence /s/+/ih/+/m/, reach to their first local maximum around frame 20. After two/three frames there is a steep amplitudes drop. The visual speech sample sets are interpreting the lip reach to its rest position approximately around 24th sample. From the 25th samples until 37th samples the second local maximum appears in the upper and lower visual speech signals while in the corner visual speech sample sets reaches to its second maximum after approximately 10 frames delay. The second local maximum, which is categorized as lip-rounding based vowels visual phoneme according to Table 5-6, is believed to be related to articulating phoneme /ax/.

$$S_3{}^U(x) = 0.6952sin(0.06948x + 0.1862) + 0.1942sin(0.5192x - 7.745) + 0.04715sin(1.008x - 6.222) + 0.1094sin(0.7301x - 5.893) + 0.1841sin(0.3017x - 2.377) \quad (5\text{-}15)$$

$$S_3{}^L(x) = 0.7447sin(0.07959x + 0.08546) + 0.1582sin(0.6547x + 1.665) + 0.1357sin(0.2675x - 1.692) + 0.05328sin(1.125x - 2.09) + 0.1035sin(0.8969x + -3.75) \quad (5\text{-}16)$$

$$S_3{}^C(x) = 1.125sin(0.07226x + 1.033) + 2.17sin(0.1912x + 5.524) + 2.439sin(0.1638x + 2.898) + 0.1381sin(0.435x + 1.188) + 0.08992sin(0.7822x + 0.1334) \quad (5\text{-}17)$$

Table 5-10: The goodness of fitting for the approximating the normalized averaged of the first 44 samples in group 3

		$S_3{}^U(x)$	$S_3{}^L(x)$	$S_3{}^C(x)$
The sum of sine functions approximation Goodness of fit	SSE	0.1198	0.07855	0.1718
	R-square	0.96	0.9761	0.9415
	Adjusted R-square	0.9385	0.9632	0.91
	RMSE	0.06788	0.05496	0.08129

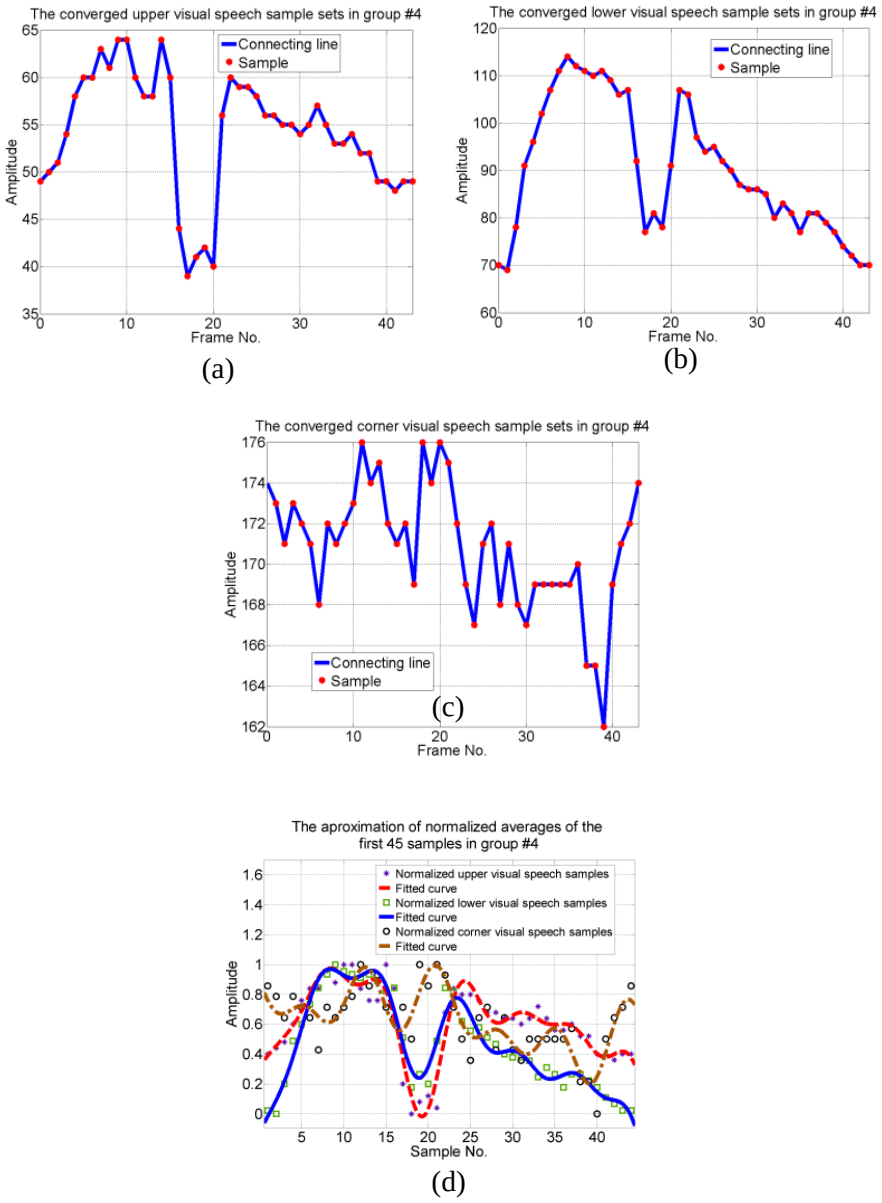


Figure 27: The (a) upper, (b) lower and (c) corner visual speech sample sets in the fourth group of the selected words and (d) the approximated polynomials

$$S_4{}^{U}(x) = 0.6631sin(0.03277x + 1.125) + 0.2269sin(0.2601x - 0.5998) + 0.1669sin(0.6511x - 1.579) + 0.1001sin(0.8604x + 0.579) + 0.2066sin(0.4236x - 3.007) \quad (5\text{-}18)$$

$$S_4{}^{L}(x) = 0.7252sin(0.07071x + 0.3812) + 0.1135sin(0.6418x + -7.378) + 0.2146sin(0.2445x - 1.004) + 0.2144sin(0.3838x - 1.892) + 0.0957sin(0.8403x - 4.892) \quad (5\text{-}19)$$

$$S_4{}^{C}(x) = 3.977sin(0.006404x + 2.791) + 0.5129sin(0.03324x + 5.186) + 0.1298sin(0.208x + 3.957) + 0.1202sin(0.5712x + 1.547) + 0.1348sin(0.8268x + 3.444) \quad (5\text{-}20)$$

Table 5-11: The goodness of fitting for the approximating the normalized averaged of the first 44 samples in the fourth group

		$S_4{}^U(x)$	$S_4{}^L(x)$	$S_4{}^C(x)$
The sum of sine functions approximation Goodness of fit	SSE	0.3244	0.1895	0.4621
	R-square	0.886	0.9547	0.7776
	Adjusted R-square	0.8309	0.9328	0.6702
	RMSE	0.1058	0.08085	0.1262

The upper visual speech sample sets has two peaks where it tends to reach to first peak after 11 samples. From 12 to 24th sample, the upper lips dimension reduces. The second peak is reached after 10 frames and finishes after 5 frames. The lower lip also has similar behaviour while in this case the 34th frame did not appear with minimum amplitude. The corner visual samples set also has two peaks approximately in the same frame numbers as upper and lower visual speech sample sets. However, the minimum value in 24th frame is higher than the other two sample sets. Therefore, lip starts articulating phoneme /s/ and /ih/ till 10th frame and closes for articulating phoneme /m/ till 24th frame. Afterward, the lip's dimension again begins to increase for articulating the rest of phonemes.

5.4. STATISTICAL ANALYSIS OF VISUAL SPEECH SAMPLE SETS

In this section, before further processing of visual speech signal sample sets (data) extracted from upper, lower and corner visual feature points, statistical analysis has been used for characterizing the statistical properties of sets. The statistical analysis includes determination of the density distribution family of the visual speech samples that they belong to as well as estimation of the density function (pdf). Determining the family of visual speech samples reveals the information about the source of data (lip's geometry) during articulating of the words' set. On the other hand, estimating the visual density function of the speech data is used in digital representation of the visual speech signals and will be discussed in Section 7.4.

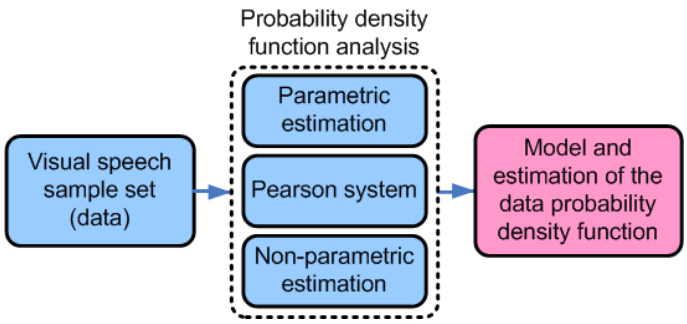


Figure 5-13: The process of analysing data density evaluation

According to Figure 5-13, the pdf analysis tool is used for determination model and probability density function of the data. The analysis is conducted using MATLAB's statistics toolbox (MATLAB, 2009).

5.4.1. PARAMETRIC DENSITY ESTIMATION

There are numbers of distributions that could be used for modelling the data's pdf. The index of choosing such pdf's is the maximum likelihood between data and estimated parameters of distributions. In this work, the parametric family of chosen data distribution consists of sixteen members. These members are Birnbaum-Saunders, Extreme value, Gamma, Generalized extreme value, Inverse Gaussian, Log-Logistic, Logistic, Lognormal, Nakagami, Negative Binomial, Normal, Poisson, Rayleigh, Rician, t location-scale and Weibull distributions. These distributions are fitted to data using the MATLAB's 'distribution GUIs' toolbox that is accessible by command 'dfittool'.

As the results of parametric pdf estimation before and after modifications show (Table 8-5) there is no suitable model that achieves the appropriate likelihood. Therefore, another type of modelling data's pdf employed that is called 'Pearson system' and will be described in Section 5.4.3.

Starting with the parametric analysis of the data distributions before and after adjusting the end points and biasing process, a set of distributions are fitted to the visual speech sample sets $u^{W_m}[n]$, $l^{W_m}[n]$ and $c^{W_m}[n]$, $n = \{0, 1, 2, \dots, F_{W_m} - 1\}$. The results show none of the fitted distribution has statistical similarity to the sample sets therefore; the Pearson system is used for finding the family of the sample sets distribution. The visual speech sample set before adjusting the end points and biasing

distribution is modelled by the parametric distributions represented in Table 8-5. The log-likelihood of fitted distributions is the index of data distribution tendency to the family of distributions. The extremely low values of the log-likelihoods are indicating the dissimilarities between the upper, lower and corner visual speech sample set distributions and the examined families of distributions.

5.4.2. NON-PARAMETRIC DENSITY ESTIMATION

The non-parametric analysis of density function (Rosenblatt, 1956) is a method of estimation the density of data distribution. The motivation of studying the non-parametric distribution estimation of data, which are containing all visual data for gaining occurrence information about sample sets amplitudes variations, is to model the upper, lower, and corner visual speech sample sets separately. Regardless to the type of visual speech sample sets distributions, the non-parametric density estimation is used to detect the actual density of the visual speech sample sets. The samples set frequencies are modelled by non-parametric density function as shown in Figure 5-14. This model reveals properties of the data source (lips) too. Such model reveals the knowledge about the source's dimensions (lip's dimensions) that has been occurring the most during pronunciation of the chosen words.

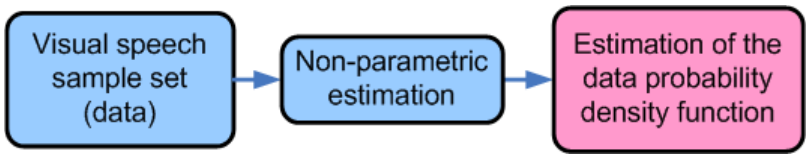


Figure 5-14: The process of determination the type of density function

This information, which is generated by the suggested interpolation method, is used particularly in digital representation of visual speech signals. Strictly speaking, in visual speech sample sets quantization scheme, the level of quantization will be defining base on the statistical characteristic of the estimated density functions of visual speech sample sets. In other words, the input quantization levels are optimized for detecting the most probable input samples and mapping them into quantized values. After analysis the TIMIT's vocabulary database and selection of a limited number of words with the most occurrence of phonemes in the first three sequential orders in the words, the visual speech sample sets have been extracted from speakers video file. The process of modification applied to visual speech sample sets resulted

in reduction of unintentional movement of head effect. Then the statistical types and distributions of them are determined. In the end of this stage, the system is ready to mutate to the next level of processing that leads to providing the main platform for introducing the concept of signature-driven visual speech analysis.

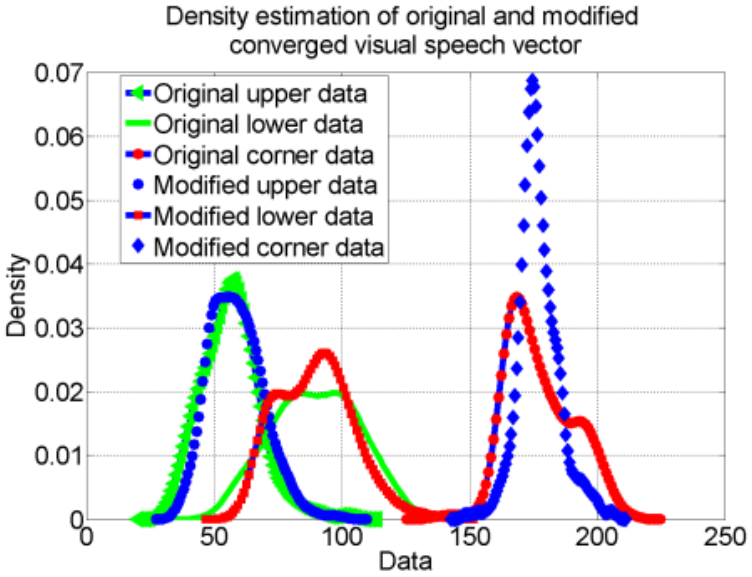


Figure 5-15: The non-parametric density estimation of the visual speech sample sets in comparison with their processed versions (endpoints modification followed by converging according to Table 5-2)

In Figure 5-15, the density functions of the original and modified upper, lower and corner data sets are illustrated by using "ksdensity" function in MATLAB. All of the estimated densities are showing bimodal distributions. The modification affects the original corner data set more than other two. This is due to imperfection in extracting pixels or the manner of pronunciation. Determination of the density family of the visual speech sample set is addressed by employing other method of density estimation. To tackle this problem, a method called Pearson system has been employed.

5.4.3. PEARSON SYSTEM

The Pearson system is a family of distributions that generalizes the differential equation of normal distribution for determining the different types of distributions. The allocated function to the Pearson system in MATLAB software is 'pearsrnd'. Using this system determines a data set distribution in absence of priori or posteriori knowledge of processed data. By solving the differential equation of normal

distribution eight possible types of distributions are obtained in such system. The type zero is the normal distribution. The other types are called as four-parameter beta distribution (type I), symmetric four-parameter beta distribution (type II), three-parameter gamma distribution (type III), not-related to any distribution (type IV), inverse-gamma location scale distribution (type V), F location scale distribution (type VI) and student's t location distribution (type VII).

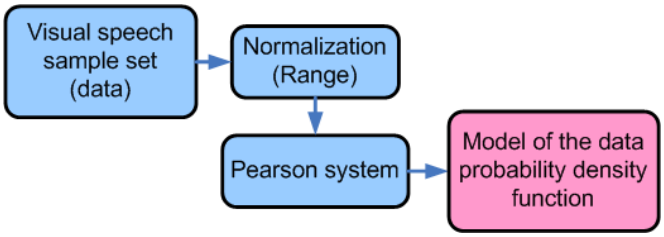


Figure 5-16: The process of determination of the model of density function in more details

Using Pearson system in the MATLAB environment as it is shown in Figure 5-16, determines the data distribution when the mean, standard deviation, skewness and kurtosis are provided. Skewness is measure of data asymmetry in probability distribution (Everitt & Skrondal, 2010) while kurtosis measures the peak of probability distribution when it is pointed or flattened (Everitt & Skrondal, 2010). These parameters are using by the Pearson function in order to model the data density function. It is revealed they all follow the beta distribution that is noted by the type I distribution in the Pearson system. Afterward, the visual speech samples distribution has been determined as the beta distribution. The Pearson system is determined the distribution family of the visual speech sample sets before adjusting the endpoints and biasing as it represented in Table 5-12.

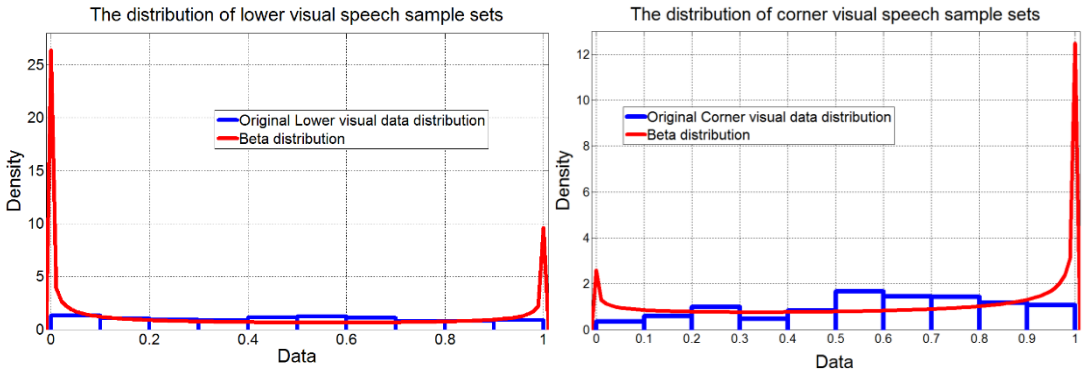
Table 5-12: The distribution family of the visual speech sample sets before adjusting the end points and biasing determined by Pearson system

	Pearson system type	
	Type	Distribution
Upper visual speech sample sets	4	Not related to any distribution
Lower visual speech sample sets	1	Beta distribution
Corner visual speech sample sets	1	Beta distribution

The Beta distribution curves are fitted to the lower and corner visual speech sample sets as it shown in Figure 5-17 (a) and (b), respectively. The Beta distribution pdf function is (MathWorks, 2008):

$$y = f(x\,|a,b) = \frac{1}{B(a,b)} x^{a-1}(1-x)^{b-1} I_{(0,1)}(x) \qquad (5\text{-}21)$$

where $B(.)$ is the Beta function and $I_{(0,1)}(x)$ is function indicator that allocates the values of x between $(0,1)$. The visual characteristics of the fitted distributions are demonstrated in Table 5-13.



(a) (b)

Figure 5-17: The probability distribution of (a) the lower and (b) corner visual speech sample sets before processing

Table 5-13: The characteristics of estimated of beta distribution before processing

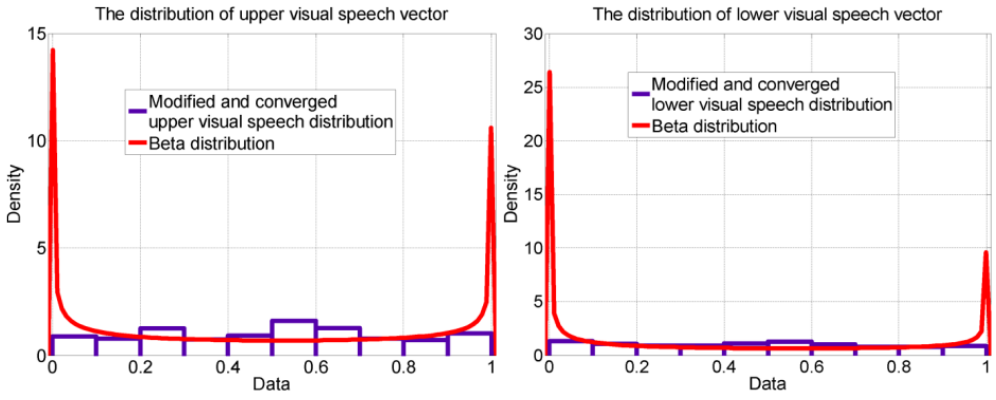
Lower visual speech signals			Corner visual speech signals		
Distribution	Beta		Distribution	Beta	
Log likelihood	57.3261		Log likelihood	62.1249	
Domain	0 < y < 1		Domain	0 < y < 1	
Mean	0.472138		Mean	0.569082	
Variance	0.108341		Variance	0.101772	
Parameter	Estimate	Std. Err.	Parameter	Estimate	Std. Err.
a	0.61395	0.0151799	a	0.802168	0.0193796
b	0.68641	0.0227049	b	0.607413	0.0184721
Estimated covariance of parameter estimates			Estimated covariance of parameter estimates		
	a	b		a	b
a	2.30E-04	7.84E-05	a	0.000375567	7.53278e-05
b	7.84E-05	0.0005155	b	7.53278e-05	0.000341219

The Pearson system determined the distribution family of the visual speech sample sets after adjusting the endpoints and biasing as it represented in Table 5-14.

Table 5-14: The distribution family of the visual speech sample sets after adjusting the end points and biasing determined by Pearson system

	Pearson system type	
	Type	Distribution
Upper visual speech sample sets	1	Beta distribution
Lower visual speech sample sets	1	Beta distribution
Corner visual speech sample sets	4	Not related to any distribution

In Figure 5-18, the beta distribution is fitted to the upper and lower visual speech sample sets where their endpoints had been modified and biased. The statistical characteristics of the fitted curves are also represented in Table 5-15.



(a) (b)

Figure 5-18: The probability distribution of (a) the lower and (b) corner visual speech sample sets after endpoints modification and converging according to Table 5-2

Comparing Table 5-12 and Table 5-14 shows a significant change of distribution for the upper and corner visual speech samples before and after modifications, respectively.

Table 5-15: The characteristics of estimated beta distribution after processing ($\boldsymbol{a}$ and $\boldsymbol{b}$ values for Eq. (5-21))

Upper visual speech signals			Lower visual speech signals		
Distribution	Beta		Distribution	Beta	
Log likelihood	95.1689		Log likelihood	162.456	
Domain	0 < y < 1		Domain	0 < y < 1	
Mean	0.484218		Mean	0.440205	
Variance	0.116384		Variance	0.120484	
Parameter	Estimate	Std. Err.	Parameter	Estimate	Std. Err.
a	0.554877	0.0120424	a	0.460139	0.0094572
b	0.591047	0.0201589	b	0.585146	0.0214023
Estimated covariance of parameter estimates			Estimated covariance of parameter estimates		
	a	b		a	b
a	1.45E-04	4.68E-05	a	8.94E-05	3.74E-05
b	4.68E-05	0.0004064	b	3.74E-05	0.0004581

5.5. LIP ANIMATION USING VISUAL SPEECH SIGNAL

A GUI (Graphical User Interface) model of lip outer contour in the MATLAB environment has been built to imitate the visual speech samples and it is shown in Figure 5-19. The dynamic of lip are rendered for preliminary observation of the visual speech samples behaviour. The motion of lip is guided by the modified visual speech sample sets (see Section 5.1)

The outer contour of lips is modelled by cubic spline curves that connect the visual speech samples. The lip contour is defined by two cubic spline curves for upper and lower contours. The upper curve and lower curves consist of seven and three nodes, respectively.

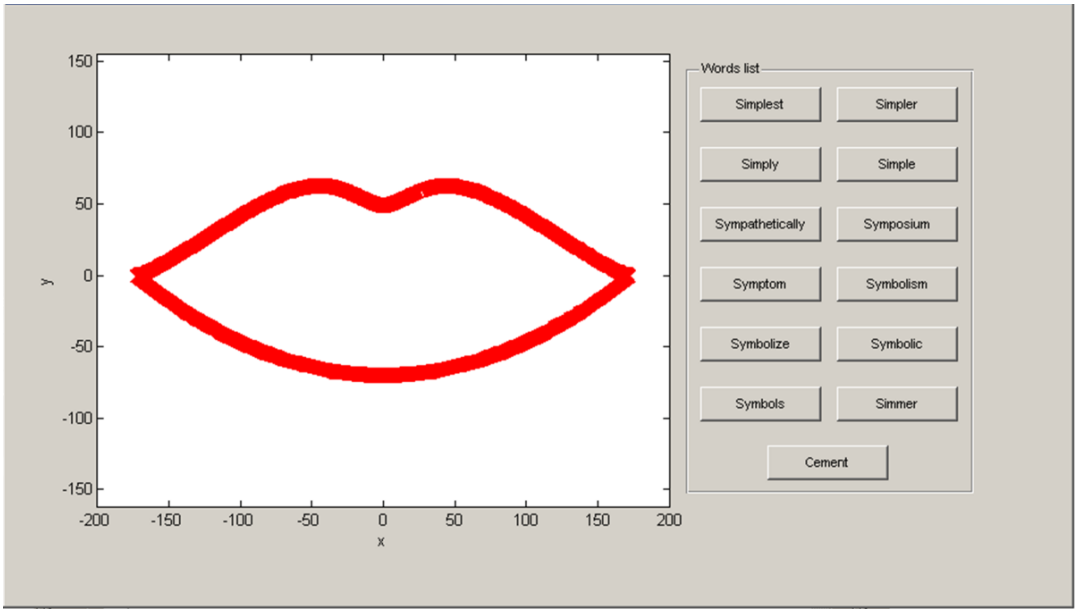


Figure 5-19: Visual speech sample set GUI

The animated lip model is used to generate more realistic observation of the movements of the mouth from visual speech sample sets during articulation of words.

5.6. SUMMARY

The extracted visual data contains the upper, lower, and corner visual speech sample sets. The sample sets inherited the effect of head movements during articulation of words. This effect in the sample sets appeared as different values in starting and stopping samples amplitudes in the first and the last visual data frames. The other issue has been the variation of starting and stopping samples amplitudes among the visual speech sample sets. The solution of the first issue had been addressed by subtracting visual speech sample sets by a line that connects the end points of each sample set. The second issue had been solved by converging the first samples to a value that had been calculated by the rounded trimmed mean of first samples from each (upper, lower and corner) visual speech sample sets. In other words, the upper, lower, and corner visual speech sample sets are biased to three constants. The next discussed topic has been the concept of normalization. In this work, the ranges of visual speech samples set are normalized for different analysis purposes. For instant, the phonemic combination /s/+/ih/+/m/ has been in common among all of words starting and observing similar trend of samples in the beginning portion of extracted visual speech sample sets has been desirable. Therefore, by using the concept of range

normalization, such unity in visual speech sample sets (visual data) would be observable. Furthermore, range normalization has been used in different section of this work such as determination of visual speech sample sets probability distribution and relating the visual speech signals. On the other hand, normalizing the domain, which has been originating from scaling the data domain, will be used in time domain representation of visual speech sample sets.

The fundamental information of the sample sets has been obtained from statistical analysis. The statistical analysis had been studied for visual speech sample sets before and after end points adjustment and convergence, (biasing) processes. The provided information included the type (model) of probability distributions and approximation of the density functions of the visual speech sample sets. The model of density of the visual speech sample sets has been determined by the Pearson system since there has been not any match in parametric models of the visual speech sample sets density functions. Using the non-parametric density estimation of visual speech sample sets reveals the shape of concentration of the most and the least probable samples. The non-parametric estimation will be used in digital representation of the visual speech signals. After processing the data and finding the fundamental statistic properties, the sample data has been ready to shape as mathematical expressions.

In the next chapter, the concept of signal construction applies to the visual speech samples. It follows by allocating a basic sample sets notations widely used in the field of signal processing theory. Finally, the performance of the BLI approach in term of its accuracy has been studied.

Chapter 6

EVALUATION OF BLI

Investigating on signal construction of visual speech data is the core of this project. Interpolation theories and discrete-to-continuous signal conversion are important to develop a mathematical expression of the visual speech signal, which have discrete-time properties. Basic functions are used in interpolation; their shifted versions are multiplied by the samples to create the constructed signals. In signal processing theory, the samples are multiplied by transfer functions similar to the interpolation process.

6.1. OVERALL VIEW OF VISUAL SPEECH PROCESSING UNIT

The visual speech signals are consist of the visual speech sample sets which produce a visual word. The identification characteristics are described as barcode allocation, digital representations, 2D/3D visual word signatures and volumetric representation of the visual words. In the signal processing approach fixed basic functions construct signals by interpolation method, regardless to the number of data samples. Figure 6-1 shows an overall view of visual speech processing unit which is the core unit in deriving the visual speech signature.

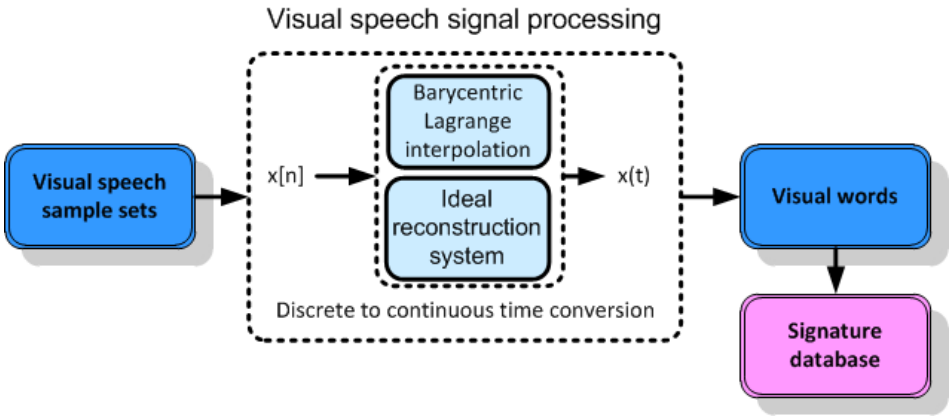


Figure 6-1: The overall view of visual speech processing unit (signal construction process)

Based on the domain property, there are two categories of the construction approach as continuous-time and discrete-time constructions. In the first approach of construction, the Barycentric Lagrange Interpolation (BLI) is used as the main construction method for deriving signatures and its performance is evaluated by the ideal construction system.

A constructed arbitrary signal using the BLI is called BL polynomial while the constructed visual speech sample sets is called visual speech signal. The visual speech samples are displayed in the discrete-time domain. Such representation tends to bridge the visual speech sample set to continuous time conversion. Before generating the visual word expressions in terms of visual speech signals, the BLI method is evaluated. The method of evaluation examines the variables that could be the subject of errors in the BLI approach.

The visual data consists of three visual speech sample sets extracted from lip's geometry during words utterance. The visual words are modelled by visual speech signals via the BLI approach so that signature allocation can operate on the visual speech signals. Thus the visual speech signals are defined as a function in the frame domain f_i. Applying the visual speech sample sets to the BLI according to Eq. (2-30), results the unique expressions of visual words in form of the mathematical representations (Section 2.12.4):

$$VS_{BLI}^{u}{}^{W_m}(f_i) = \frac{\sum_{i=0}^{F_{W_m}-1} u^{W_m}[i] \frac{w_i}{(f_i - i)}}{\sum_{j=0}^{F_{W_m}-1} \frac{w_i}{(f_i - i)}} \tag{6-1}$$

$$VS_{BLI}^{l}{}^{W_m}(f_i) = \frac{\sum_{i=0}^{F_{W_m}-1} l^{W_m}[i] \frac{w_i}{(f_i - i)}}{\sum_{j=0}^{F_{W_m}-1} \frac{w_i}{(f_i - i)}} \quad (6\text{-}2)$$

$$VS_{BLI}^{c}{}^{W_m}(f_i) = \frac{\sum_{i=0}^{F_{W_m}-1} c^{W_m}[i] \frac{w_i}{(f_i - i)}}{\sum_{i=0}^{F_{W_m}-1} \frac{w_i}{(f_i - i)}} \quad (6\text{-}3)$$

where the number of samples in interpolating signal is $f_i = \beta F_{W_m}$. The β refers to a desirable number of spans and equivalent to the sampling frequency, where $f_i \in [0, \beta(F_{W_m} - 1)]$, $m = 1, 2, 3, \dots, w$. As β reaches the infinity, the polynomials get to the ideal continues state. In this study, visual speech signal are constructed by $\beta = 150$. It represents the continuous state of polynomial, adequately. The normalized ratio of upper to corner and lower to corner visual speech sample sets are denoted by ${N_{uc}}^{W_m}[n]$ and ${N_{lc}}^{W_m}[n]$, where $i \in [0: (F_{W_m} - 1)]$ and $m = 1, 2, 3, \dots, w$, respectively.

$${N_{uc}}^{W_m}[n] = \sum_{k=0}^{F_{W_m}-1} \left(\frac{u_i^{W_m}[k]}{c_i^{W_m}[k]}\right) \delta[n-k] \quad (6\text{-}4)$$

$${N_{lc}}^{W_m}[n] = \sum_{k=0}^{F_{W_m}-1} \left(\frac{l_i^{W_m}[k]}{c_i^{W_m}[k]}\right) \delta[n-k] \quad (6\text{-}5)$$

The visual speech signals corresponding to the sample sets in Eq. (6-4) and (6-5) are:

$$VS_{BLI}^{uc}{}^{W_m}(f_i) = \frac{\sum_{i=0}^{F_{W_m}-1} {N_{uc}}^{W_m}[i] \frac{w_i}{(f_i - i)}}{\sum_{i=0}^{F_{W_m}-1} \frac{w_i}{(f_i - i)}} \quad (6\text{-}6)$$

$$VS_{BLI}^{lc}{}^{W_m}(f_i) = \frac{\sum_{i=0}^{F_{W_m}-1} {N_{lc}}^{W_m}[i], \frac{w_j}{(f_i - i)}}{\sum_{i=0}^{F_{W_m}-1} \frac{w_j}{(f_i - i)}} \quad (6\text{-}7)$$

The concatenated normalized ratio of upper to corner ${N_{uc}}^{W_m}[n]$ and lower to corner visual speech sample sets ${N_{lc}}^{W_m}[n]$, is denoted by $uclc_i^{W_m}[i]$:

$$uclc^{W_m}[n] = \sum_{k=0}^{2F_{W_m}-1} uclc_i^{W_m}[k] \delta[n-k] \quad (6\text{-}8)$$

The notation $uclc_i^{W_m}[i]$ represents upper, lower and corner concatenation of visual speech sample sets. Eq. (6-8) is also applied to the BLI formula (Eq. (2-30)) resulting in representation of following visual words:

$$VS_{BLI}^{uclc^{W_m}}(f_i) = \frac{\sum_{i=0}^{2F_{W_m}-1} uclc^{W_m}[i]\frac{w_i}{(f_i - i)}}{\sum_{i=0}^{F_{W_m}-1}\frac{w_i}{(f_i - i)}} \quad (6\text{-}9)$$

As mentioned, the mathematical expression driven by Eq. (6-9) cannot be produced by MATLAB. Therefore, the numerical representations of concatenated visual speech signals are available. Therefore, the concatenated visual speech signal can be referred to as unique numerical representation of each visual word. Besides, using the BLI provides the ability of frame rate exchange to 60^{fps} or 120^{fps} by simply interpolating the visual speech samples by $\beta = 2$ and $\beta = 3$, respectively. It is due to the fact that capturing the lip movements is performed by 30^{fps}. The mathematical expressions are represented in Appendix D in Eq. (8-1) to (8-65) comprehensively.

In the BLI, it is expected to increase the similarity of a function and interpolating polynomial by increasing the number of associated samples from the function. As shown previously, the Lagrange Interpolation did not satisfy the expectation and did not diagnose with the Runge's effect. The Runge's effect is eliminated by both basic function modifications and the Chebyshev node insertion which led to BLI. In the BLI, like other interpolation methods, the number of samples plays an important role in goodness of interpolation. Therefore, the BLIs of two types of sample sets are evaluated in terms of the number of samples and amplitudes provided by each sample set. In addition to these factors, the BLI demonstrates a phenomenon proportional to the rate of its sampling frequency β.

It is shown that based on the BLI; there is a direct relation between the accuracy of extracting samples back from interpolating polynomial and the sampling frequency. As shown, the BLI method is not exactly include the amplitudes in the resulted polynomial. There is always an error, which is not eliminated even by increasing the sampling frequency. Therefore, the evaluation has been studied. The evaluation has been conducted based on convergence rate of the BLI. Besides, the BLI convergence rate is also determined by two factors as the number of samples and the amplitude of

samples. It means how well BLI interpolates a function with fixed amplitudes while the numbers of samples are variable and vice versa. More specifically, construction error can be determined by alternating the length of samples (number of samples) or the sample set's amplitudes. It has been assumed that the sample sets are represented by $f(x)$. In the first scenario, the amplitude is kept fixed; as a result, the set construction associated with variable sample lengths determines the difference between the function $f(x)$ and its interpolation denoted by $\tilde{f}(x)$.

In the second scenario, the evaluation is arranged by changing the amplitude of the sample sets while the length of samples is assumed fixed. A function for using both evaluation methods is chosen introduced by Gaussian curve. In addition, to gain more realistic predictions form the outcome of visual speech signals, the extracted visual speech samples are averaged. Then their approximated polynomials are used to evaluate the performance of BLI under the category of deterministic signals. Such process of interpolation can provide geometrical similarity between constructed signal and the original signal when a limited number of samples often less than the original signal have been provided.

The other important issue is the dissimilarity between the original polynomial $f(x)$ and the interpolating polynomial $\tilde{f}(x)$ caused by varying the number of samples. Therefore, it is very important to be aware of both the interpolating polynomial signal and the original signal sizes otherwise comparison is not possible. The similarity of an interpolating polynomial $\tilde{f}(x)$ subjected to $f(x)$ based on the interval $[a: b]$ and density function ρ is defined by a theorem called accuracy of the polynomial interpolation (Trefethen, 2000). It states where the interpolation points $N \to \infty$, there is a constant $0 < c$ for all values of $x \in [a: b]$ that satisfies the amount of error by:

$$\left|f(x) - \tilde{f}(x)\right| \leq e(x) \tag{6-10}$$

The numbers of nodes are identical to the number of equidistance spans with the length equal to one. The error function also represents a peak on the residual interval. Instead of all sample errors, the peak value on the residual considers as an indication for error determination. Therefore, the Eq. (6-10) can be rewritten as:

$$max\left|f(x) - \tilde{f}_j(x)\right| \leq ce^{-N} \tag{6-11}$$

for $j = \{0,1,2,\dots,k\}$ where k is the number of various interpolating functions. The interpolating functions $\tilde{f}_i(x)$ are defined as different versions of the deterministic function $f(x)$. Considering Eq. (6-11) it is noted whenever a function $f(x)$ is interpolating by N points from $\tilde{f}_i(x)$, the maximum residual of them should be less than or equal to a decaying exponential function scaled by a positive constant c. Therefore, the left side of Eq. (6-11) is a sequence of errors for each node. In order to examine the accuracy of signal reconstruction, the sample sets dimensions (amplitude and sampling rate) are considered.

6.2. RECONSTRUCTION WITH VARIABLE LENGTHS AND FIXED AMPLITUDES

The interpolating polynomials are generated by changing the sampling frequency of function $f(x)$. As shown in Figure 6-2 the reconstruction input is defined as deterministic function source consisting of both Gaussian function and the approximated polynomials on a portion of normalized average of the upper, lower and corner visual speech samples.

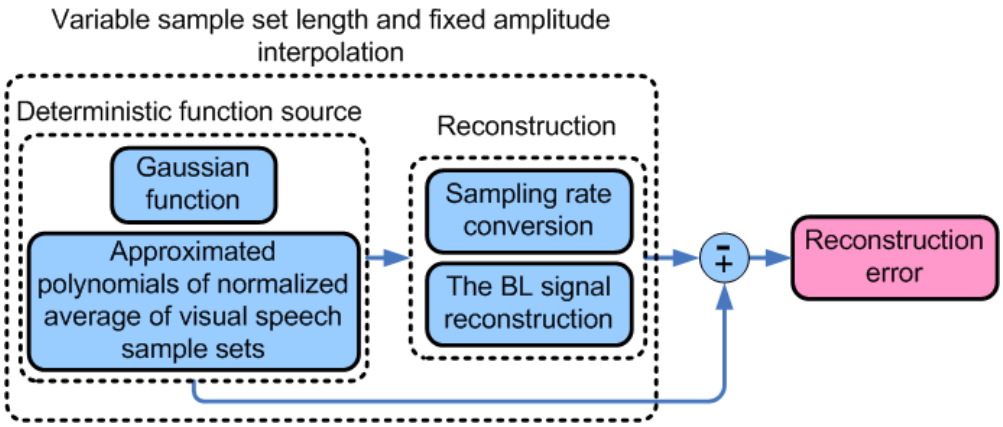


Figure 6-2: The method for evaluating the reconstruction error for the BLI approach (variable sample set length, fixed amplitude)

The samples are taken from $\beta - 1$ equidistance spans based on the interval $[a: b]$ with $N - 1$ different lengths is determined from $f(x)$. Each span is defined as follow:

$$x = a + i\frac{b-a}{N-1} \tag{6-12}$$

where $i = \{1, 2, 3, \dots, N-1\}$. Based on Eq. (6-12) the matrix form can be assumed as an upper or lower triangular matrix. By substituting these samples into $f(x)$, the $N-1$ versions are defined. Afterward, these functions are ready to be interpolated with the BLI with a fixed number β. It yields interpolating polynomial functions $\tilde{f}(x)$, $i = \{2, 3, 4, \dots, N-1\}$ with equal length of β samples. Note that, the first entity of interval cannot be used since it consists of a single value. In this stage, it is possible to compare $f(x)$, which also defined by β with $\tilde{f}(x)$ polynomial. This approach reveals the degree of similarities between $f(x)$ and its constructed versions as the number of available samples for reconstruction is increasing. It should be mentioned that the BLI algorithm in this study has been modified to include the samples by replacing the NaN (Not A Number) values in standard IEEE arithmetic caused by dividing the basic functions by zero on $x \in [0, F_{W_m} - 1]$. In addition, the maximum number of samples, which could be used in this algorithm, is limited to $N = 169$ samples.

By using curve fitting toolbox in MATLAB, the deterministic functions are represented in Eq. (6-13) to (6-15) for error evaluation of the BLI, with variable lengths and fixed amplitudes:

$$\begin{aligned} f_u(x) = {} & 0.003824exp(-((x-11.01)/0.0298)\text{^}2) + 0.9462exp(-((x \\ & -9.624)/4.504)\text{^}2) + 0.6264exp(-((x \\ & -36.59)/5.051)\text{^}2) + 0.3006\,exp(-((x-32.01)/3.74)\text{^}2) \\ & + 0.4933\,exp(-((x-16.71)/4.799)\text{^}2) \end{aligned} \tag{6-13}$$

$$\begin{aligned} f_l(x) = {} & 0.8413exp(-((x-11.54)/4.068)\text{^}2) + 0.4982\,exp(-((x \\ & -7.204)/2.921)\text{^}2) + 0.119\,exp(-((x \\ & -16.21)/2.268)\text{^}2) + 0.7008\,exp(-((x \\ & -34.11)/8.054)\text{^}2) + 0.5784\,exp(-((x \\ & -19.79)/6.136)\text{^}2) \end{aligned} \tag{6-14}$$

$$\begin{aligned} f_c(x) = {} & 0.7345\,exp\,(-((x-17.91)/4.131)\text{^}2) + 0.8708exp(-((x \\ & -11.5)/4.777)\text{^}2) + 0.1996exp(-((x-22.67)/c3)\text{^}2) \\ & + 0.3288exp(-((x-1.947)/2.731)\text{^}2) + 0.6475exp(-((x \\ & -36.42)/6.4)\text{^}2) \end{aligned} \tag{6-15}$$

Having a sampling frequency of 8000 samples, these approximated polynomials are shown in Figure 6-3 (a). The goodness of polynomial fitting is represented Table 6-1. The interpolating polynomials are identified by $\tilde{f}_u(x)$, $\tilde{f}_l(x)$ and $\tilde{f}_c(x)$.

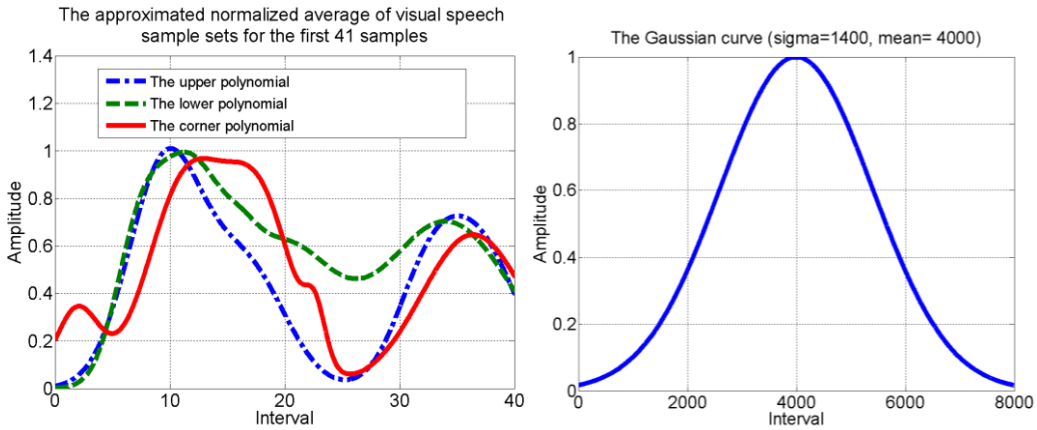
Table 6-1: The goodness of fit in approximating the polynomials $\boldsymbol{f_u(x)}$, $\boldsymbol{f_l(x)}$ and $\boldsymbol{f_c(x)}$

	Goodness of error fitting			
	$f_u(x)$	$f_l(x)$	$f_c(x)$	$f_u(x)$
SSE	0.02417	0.007435	0.04121	0.02417
R-square	0.9934	0.9968	0.9875	0.9934
Adjusted R-square	0.9883	0.9951	0.9805	0.9883
RMSE	0.03315	0.01725	0.0406	0.03315

The Gaussian curve is defined by the MATLAB function:

$$f(x) = gaussmf(x, [sig, c]) \tag{6-16}$$

where $sig = 1400$ and $c = 4000$ with sampling frequency 8000 samples are selected for continuous state of functions. It should be mentioned that the value of β is variable as shown in Figure 6-3 (b).



(a) (b)

Figure 6-3: The approximated (a) visual speech polynomials $\boldsymbol{f_l(x)}$, $\boldsymbol{f_l(x)}$ and $\boldsymbol{f_c(x)}$ (b) and Gaussian curve $\boldsymbol{f(x)}$

The interpolating polynomial is identified by $\tilde{f}_g(x)$. The errors are illustrated in Figure 6-4 and are approximated with exponential functions in Eq. (6-13) to Eq. (6-15) for the upper, lower, corner and Gaussian curves denoted

by $e_{ux}(x)$, $e_{lx}(x)$, $e_{cx}(x)$ and $e_{gx}(x)$, respectively. The results indicate a reduction in error rate with respect to increase in the number of samples.

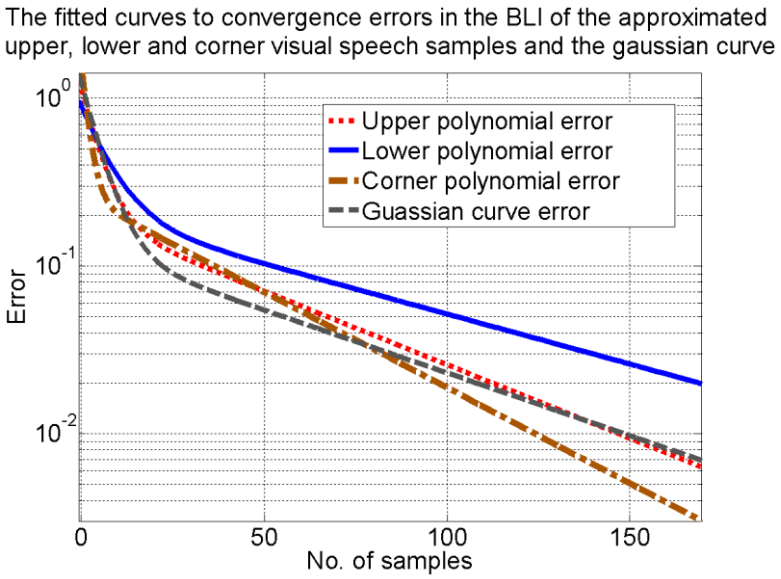


Figure 6-4: The error curves in the BLI approximation of the $\boldsymbol{f(x)}$, $\boldsymbol{f_u(x)}$, $\boldsymbol{f_l(x)}$ and $\boldsymbol{f_c(x)}$ with variable number of samples and fixed amplitudes

$$e_{ux}(x) = 0.9412exp(-0.2235x) + 0.1936exp(-0.02017x) \quad (6\text{-}17)$$

$$e_{lx}(x) = 1.345exp(-0.5466x) + 0.2598 * exp(-0.02626x) \quad (6\text{-}18)$$

$$e_{cx}(x) = 0.6928exp(-0.1441x) + 0.2039exp(-0.01373x) \quad (6\text{-}19)$$

$$e_{gx}(x) = 1.111exp(-0.2053x) + 0.1291exp(-0.01723x) \quad (6\text{-}20)$$

The goodness of curve fittings with exponential function is shown in Table 6-2.

Table 6-2: The goodness of curve fitting to $\boldsymbol{e_{ux}(x)}$,$\boldsymbol{e_{lx}(x)}$,$\boldsymbol{e_{cx}(x)}$ and $\boldsymbol{e_{gx}(x)}$

	Goodness of error fitting			
	$e_{ux}(x)$	$e_{lx}(x)$	$e_{cx}(x)$	$e_{gx}(x)$
SSE	0.1009	0.06897	0.06643	0.03813
R-square	0.9646	0.9691	0.9757	0.9882
Adjusted R-square	0.9639	0.9685	0.9753	0.988
RMSE	0.02481	0.02051	0.02013	0.01525

6.3. RECONSTRUCTION WITH VARIABLE AMPLITUDES AND FIXED LENGTHS

The variation of amplitude of $f(x)$, which could affect the interpolation error, is the other factor. According to Figure 6-7, this time the length of samples are fixed while the amplitude of the function is changing. Similar to the reconstruction of sample set with variable length, $f(x)$ in this scenario consists of different amplitude versions.

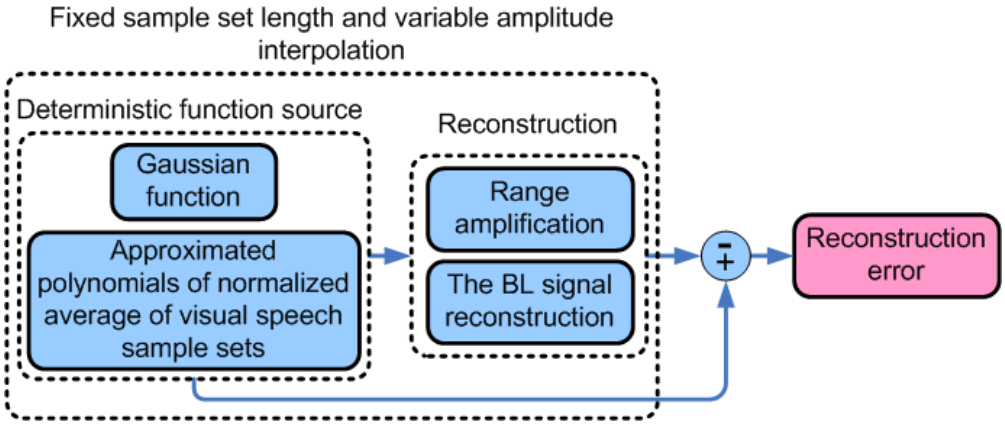


Figure 6-5: The method for evaluating the reconstruction error for the BLI approach (fixed sample set length, variable amplitude)

The BLI is using to determine the interpolating polynomial with a constant length. The main function $f(x)$ is defined with sampling frequency of samples n called $f'(x)$. In parallel, the same function is defined by sampling frequency of βn samples called $f''(x)$. The different amplified versions are obtained for both functions by simply multiplying $f'(x)$ and $f''^{(x)}$ by a range between $[m{:}\,h]$where m equals to the peak of each function and h refers to the number of amplified versions. The interpolating results are denoted by $f_h{}'(x)$ and $f_h{}''(x)$ for $h = \{m, m+1, m+2, \dots, hh$. In the final stage, the amplified functions in $f_h{}'(x)$ are interpolated by sampling frequency βn. The error is determined by maximum difference between these interpolating polynomials and the ones in $f_h{}''(x)$. Unlike the pervious case, the error rate is increasing according to the obtained results.

After fixing the number of sample sets and varying the amplitudes of them, the second approach BLI error evaluation should be employed. In Figure 6-6, the amplitudes in Eq. (6-13) to (6-15) change between $m = 1$ and $h = 300$ while the

number of samples is fixed to $n = 40$ samples. The Gaussian function is characterized by $sig = 6$ and $c = 20$. The sampling frequency is set to $\beta = 200$.

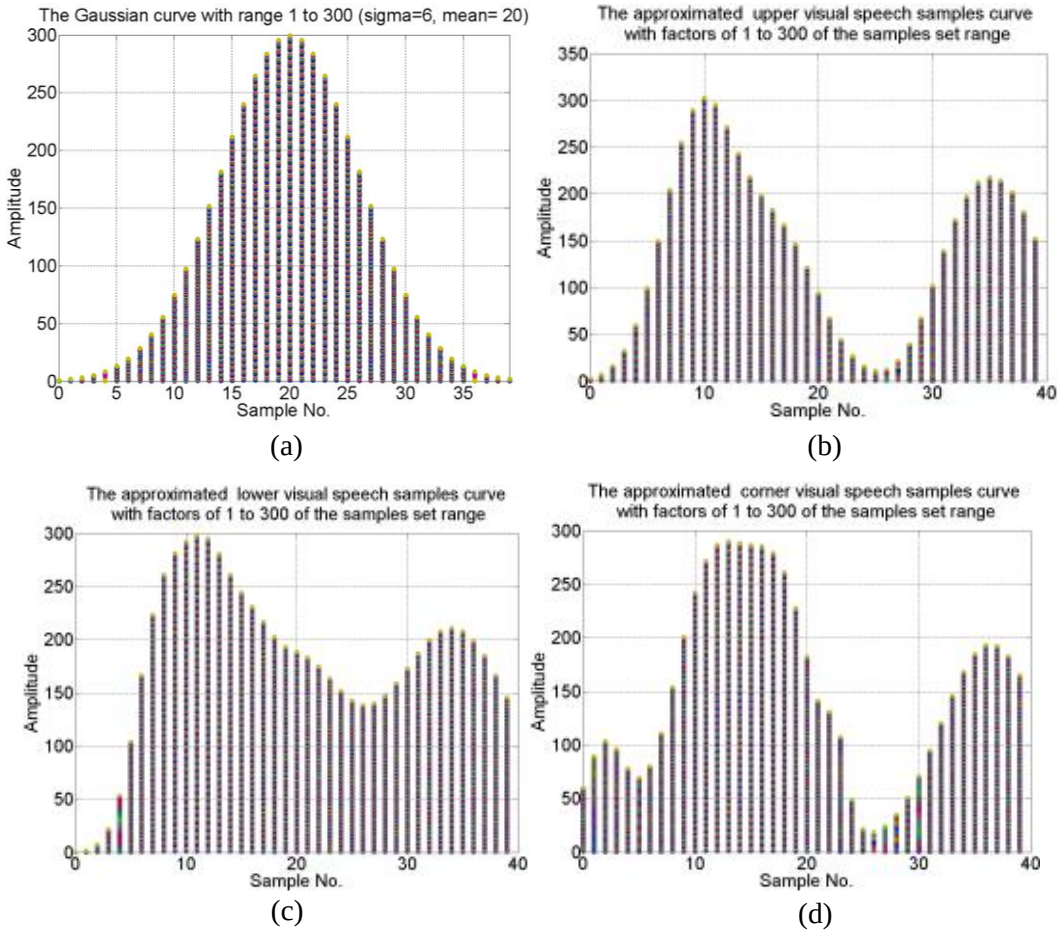


Figure 6-6: The provided sample sets from (a) Gaussian curve first 40 samples of a approximated polynomial from normalized average of (b) upper, (c) lower, (d) corner visual speech sample sets with amplitude variation between 1 and 300

In Figure 6-7, the errors are illustrated by fitting exponential functions that are represented in Eq. (6-21) to Eq. (6-24) for the upper, lower, corner and Gaussian curves denoted by $e_{uy}(x)$, $e_{ly}(x)$, $e_{cy}(x)$ and $e_{gy}(x)$, respectively.

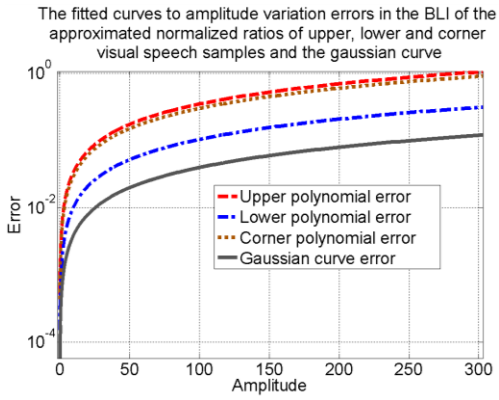


Figure 6-7: The BLI approximation errors obtained by fixed number of samples and variable amplitude scheme

$$u(x) = 0.003467x\ - 3.087e - 015 \tag{6-21}$$

$$l(x) = \ 0.001032x\ +\ 2.051e - 015 \tag{6-22}$$

$$c(x) = \ 0.002986\,x\ +\ 1.872e - 016 \tag{6-23}$$

$$g(x) = \ 0.0003949x\ +\ 7.265e - 016 \tag{6-24}$$

Table 6-3: The goodness of curve fitting to $\boldsymbol{e_{uy}(x)}$,$\boldsymbol{e_{ly}(x)}$,$\boldsymbol{e_{cy}(x)}$ and $\boldsymbol{e_{gy}(x)}$

	Goodness of error fitting			
	$e_{uy}(x)$	$e_{ly}(x)$	$e_{cy}(x)$	$e_{gy}(x)$
SSE	6.29E-25	3.30E-25	3.03E-26	4.14E-25
R-square	1	1	1	1
Adjusted R-square	1	1	1	1
RMSE	4.59E-14	3.33E-14	1.01E-14	3.73E-14

As the amplitudes are increasing, the error curves also increase as it is observed from Figure 6-7. The maximum error, which belongs to the upper polynomial, is equal to 1 when reached by changing the amplitudes.

6.4. SUMMARY

In this chapter, the performance of BLI method is evaluated using two different aspects. First, the construction of sample sets when the number of samples is variable and their amplitudes is fixed, and the second one, when the amplitudes are variable and the numbers of samples are assumed fixed. The results indicate that the accuracy

of the BLI increases as the number of provided samples from a function increases. Also, the accuracy reduces while the amplitudes of samples increase. Moreover, evaluating the BLI is performed by affecting the sampling frequency. It has been shown that, an expression of visual speech signal does not satisfy the recovering of the lip's movement if a sampling frequency be less than a certain value.

It has been revealed that the accuracy of BLI in reconstruction of signals is increased while the amplitudes of samples kept constant and the numbers of provided samples are increased. On the other hand, the accuracy of BLI is reduced in case of increasing the amplitudes of samples when the numbers of samples are constant. Therefore, it can be suggested that the BLI would have better performance if the amplitudes of sample sets are not scaled.

Chapter 7

ALLOCATING VISUAL WORD SIGNATURES

In this chapter, the process of building different variations of visual words and their signatures has been described. The signatures have been categorised as 2D/3D visual words, visual words' barcode, digital visual word and volumetric representation of visual words. Allocating explicit expression to the visual data is the first step toward defining visual words signature. Furthermore, the performance of recovering the visual speech sample sets from BLI approach will be examined. In addition, finding their relations for more simplification, reducing their dimension allocate identification properties to the visual words. In Figure 7-1, the process of signature allocation to the visual words has been demonstrated.

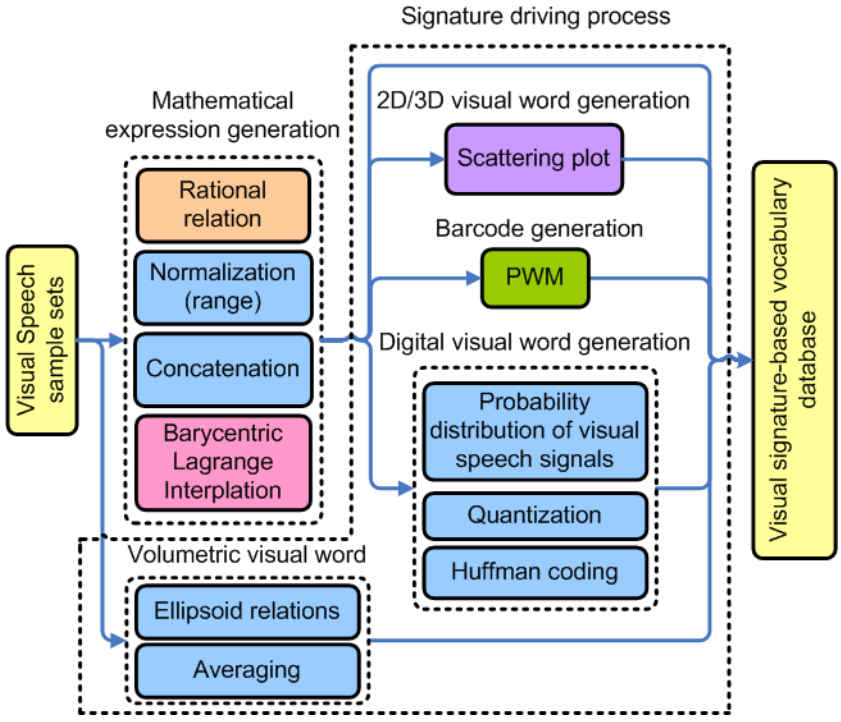


Figure 7-1: The overall view of signature allocation to the visual speech sample sets

In this thesis, the mathematical expression generation unit is the fundamental core of the methodology for analysing the visual speech data. As mentioned, the extracted samples from a visual feature are called visual speech sample set. Collection of three visual speech sample sets corresponding to the upper, lower and corner visual feature points is representing a word.

7.1. VISUAL SPEECH SIGNAL REPRESENTATIONS

The visual speech signals can be stored by their mathematical expressions. However, such collection does not present efficiency in terms of storage, readability, and simplicity. Therefore, it is desirable to consider some of the important characteristics of an abstract database. Typically, database is a collection of information. The emphasis in this part is on the arrangement of information. Therefore, creating a database should follow a set of specifications. The database members should have a rational simplicity after creation while maintaining the ability of recovering the visual speech sample sets before processing. The simplicity of database's categories can cover the necessity of memory space minimization resulting in computational cost reduction. The other important characteristic of a desirable database is unification of the stored data. The lossless storage's processing is the other important dimension that needs to be considered. In other words, all the database's members need to be recovered completely and perfectly. The database categories also need to be unified and standardized. For example, three visual speech sample sets should be simplified and be represented in a single expression. In this way, the database has the combination of expressions and signatures. Expressing the visual speech signals of each word has been calculated although in case of deriving visual words database a simple storage of visual word's signals could not satisfy the criteria of minimized memory usage and compactness. For this reason, the other methods of visual words signature extraction, which are accommodating the efficiency to the visual speech signals signatures, are proposed.

The visual speech signals are constructed by applying the BLI to these three sample sets are called visual word (see Figure 7-2). After relating the visual speech sample sets by their ratios (rational expressions) and normalizing them, further dimensional reduction on the visual speech signals leads to the concatenating them. The results of mathematical expression generation unit are used for signature driven process for the

visual words. Concatenation of two normalized ratios of sample sets is led to a unique numerical representation of each visual word. The main issue in this form of visual word representation is the lack of mathematical expressions due to the employed software computational restrictions (The message 'Output truncated. Text exceeds maximum line length of 25,000 characters for Command Window display' appears in the MATLAB command window). The visual speech signal from normalized ratio the upper, lower and corner visual feature samples are more simplified versions of the visual words. In this level, each visual word is expressed by two visual speech signal called visual word's signature expressions caused reduction in number of polynomials. From this level forward, the visual word's signatures are processed with different methods to derive visual word barcode, 2D/3D visual word representation, digital visual word and volumetric representation of the visual words.

The visual word barcode can be used for identification of the visual word signatures and the digital visual word can approximate the visual word signatures since there are some information that has been lost during quantization and coding processes.

The ability of recovering the visual speech sample sets has been facilitated perfectly by the visual word signature and 2D/3D visual word signature. In this stage, one of the significant results of this thesis is introduced by 2D visualization of the visual speech signals constructed from the ratio of the upper to corner and lower to corner visual speech sample sets. This representation uniquely defines visual speech signature for visual words.

For volumetric representation of the visual words, the upper, lower, and corner visual features are related via three possible combination ellipses. These three ellipses are averaged and lead to a single ellipse that models the feature points and consequently define a lip template.

The result of such representations and expressions along with mathematical expressions derived for both visual speech signals and visual word's signatures are stored in a database called visual signature-based vocabulary database. The arrangements of components are described in this section. It should be noted that the BLI of upper, lower and corner sample sets without normalization or rational

expression have been stored to the database. These signals are called upper, lower corner visual speech signals.

These four methods are describing visual words by including different kind of information. The word collection has 13 members and each of word is represented by three visual speech signals. Therefore, the number of visual words in mathematical form in database is equal to 39.

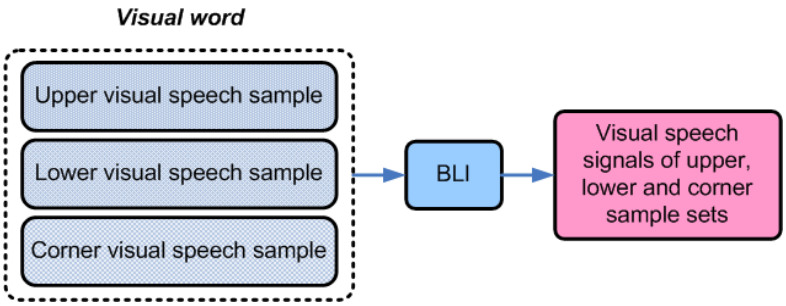


Figure 7-2: Allocating three visual speech signals to a visual speech sample sets

The mathematical expressions of visual speech signals are referred to as visual words. For instant, the mathematical expression for the upper visual speech sample set in word “Simplest" is derived by MATLAB symbolic toolbox and represented by Eq. (8-1) Appendix E.

The visual speech signals ${VS_{BLI}^{u}}^{W_m}(f_i)$, ${VS_{BLI}^{l}}^{W_m}(f_i)$ and ${VS_{BLI}^{c}}^{W_m}(f_i)$, $m = \{1,2,3,\dots,13\}$ have similar pattern to those shown in Figure 4-14, but this time they are formulated by the mathematical expressions. In Figure 7-3, all the visual speech signals except the words ‘symbolism’ and ‘cement’ reach to a minimum where it is occurred between 21st and 30th frames. This fact indicates the closure of upper lip to a minimum for finishing articulation of the phoneme /m/. Although reaching to a minimum is moderated, a similar effect is also appearing for the lower and corner visual speech signals except the words ‘symbolism’ and ‘symbolize’ where the minimums are less than 20th frame as it shown in Figure 7-4 and Figure 7-5.

- **UPPER VISUAL SPEECH SIGNALS**

The visual speech signals ${VS_{BLI}^{u}}^{W_m}(f_i)$, $m = \{1,2,3,\dots,13\}$ are shown in Figure 7-3.

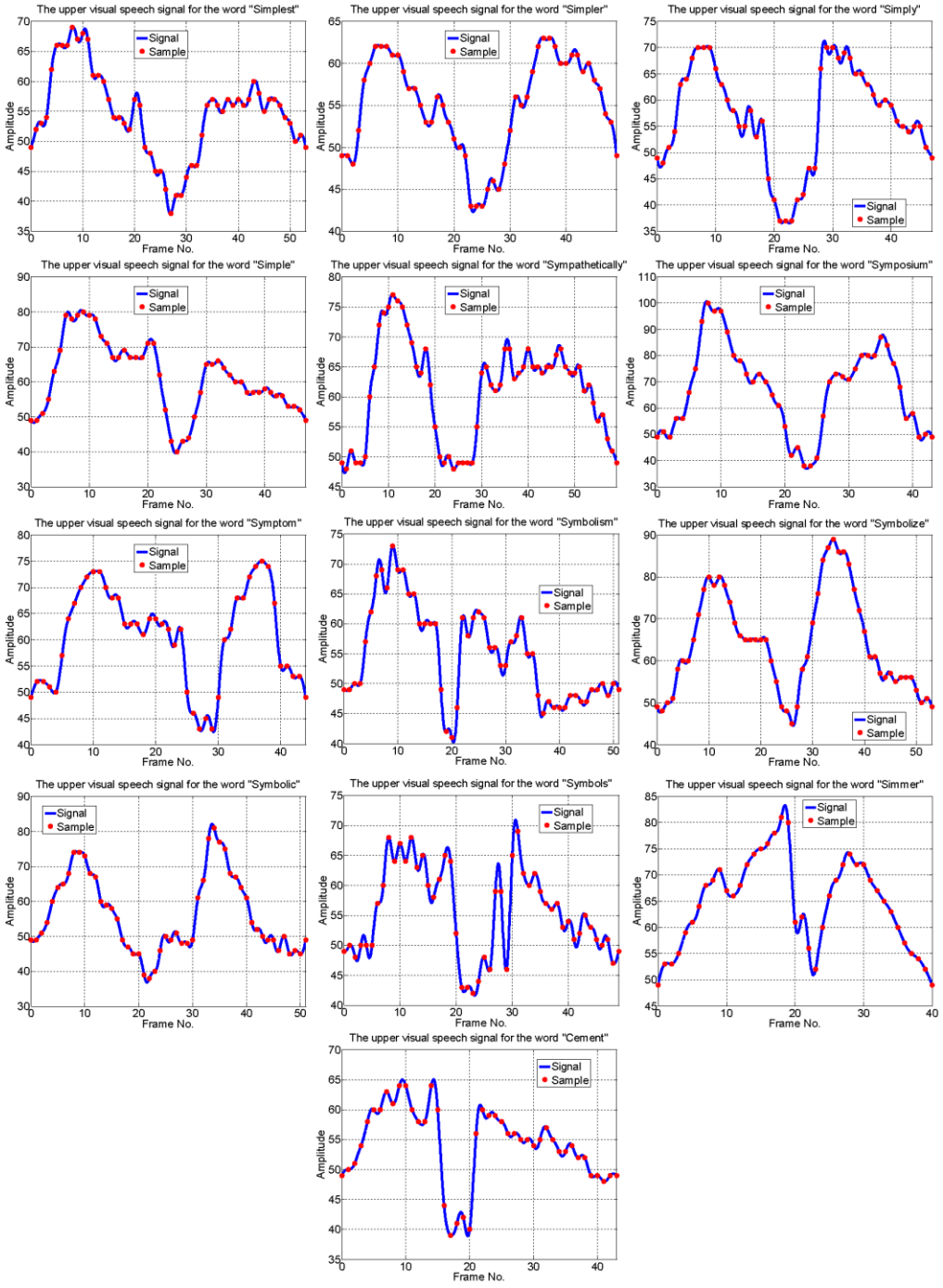


Figure 7-3: The upper visual speech signals of the selected words

- **LOWER VISUAL SPEECH SIGNALS**

The visual speech signals ${VS_{BLI}^{l}}^{W_m}(f_i), m = \{1,2,3,\dots,13\}$ are shown in Figure 7-4.

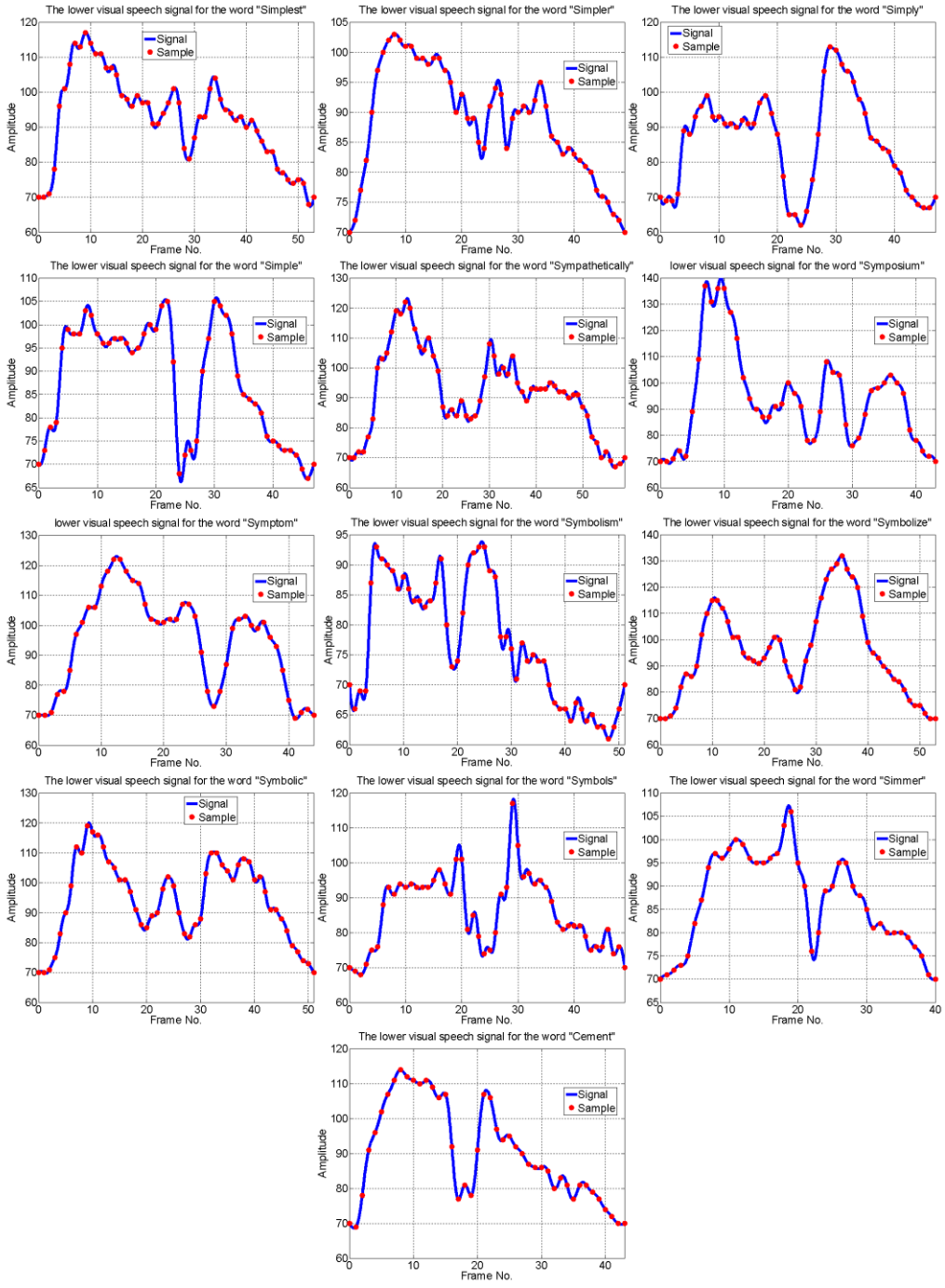


Figure 7-4: The lower visual speech signals of the selected words

- **CORNER VISUAL SPEECH SIGNALS**

The visual speech signals$VS_{BLI}^{c}{}^{W_m}(f_i), m = \{1,2,3,\dots,13\}$ are shown in Figure 7-5.

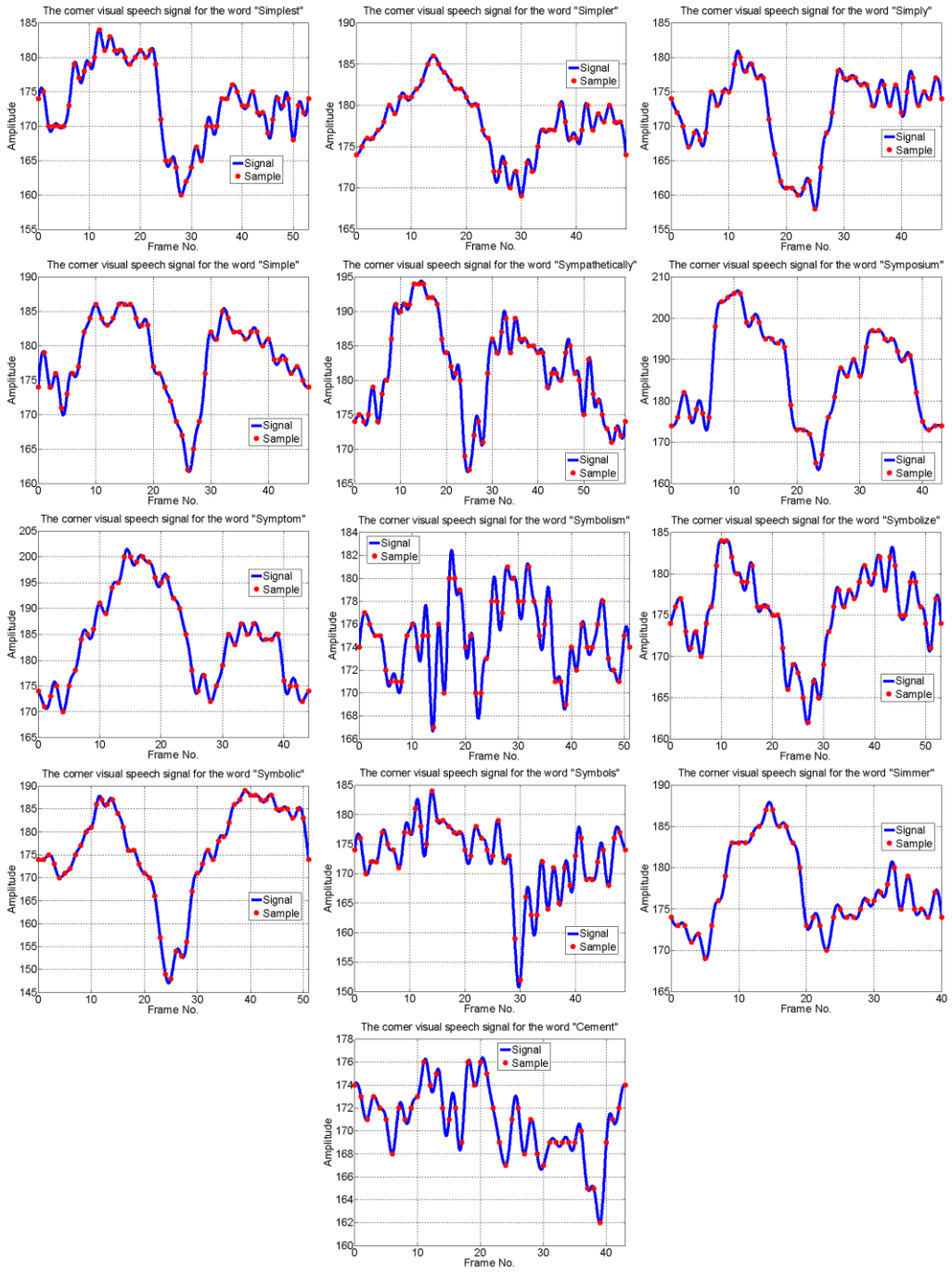


Figure 7-5: The corner visual speech signals of the selected words

The overall representations of the visual words, which are separated according to the three visual features, are depicted for the upper, lower and corner visual speech signals in Figure 7-6 (a) to (c).

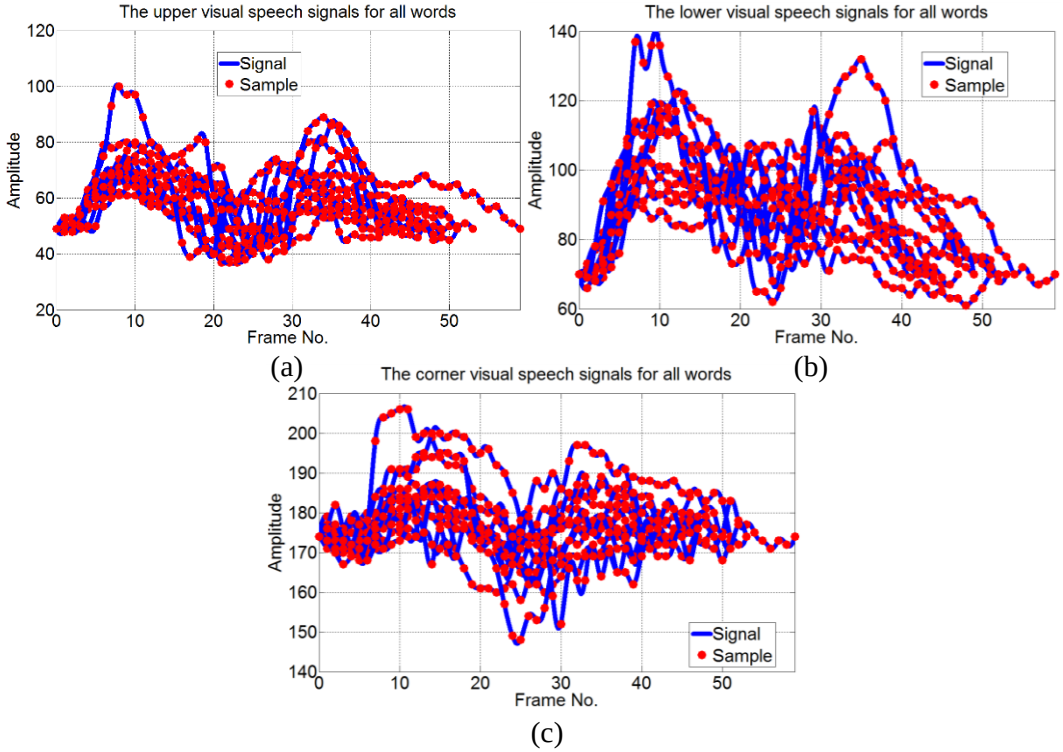


Figure 7-6: The visual speech signals of (a) the upper, (b) lower and (c) corner visual speech sample sets

Based on the necessities for categorizing data, the process of creating signature expression uses two complementary components that would be combined with the construction method (BLI). For creating a standardized database, the process of normalization has been employed. Relating the visual speech signals together in each visual word is obtained using the ratios among them. Therefore, compactness of expressions has been satisfied while the lossy compression avoided. All the possible combinations (configurations) of construction method, normalization process and ratio calculation for obtaining the optimized database is shown in Figure 7-7.

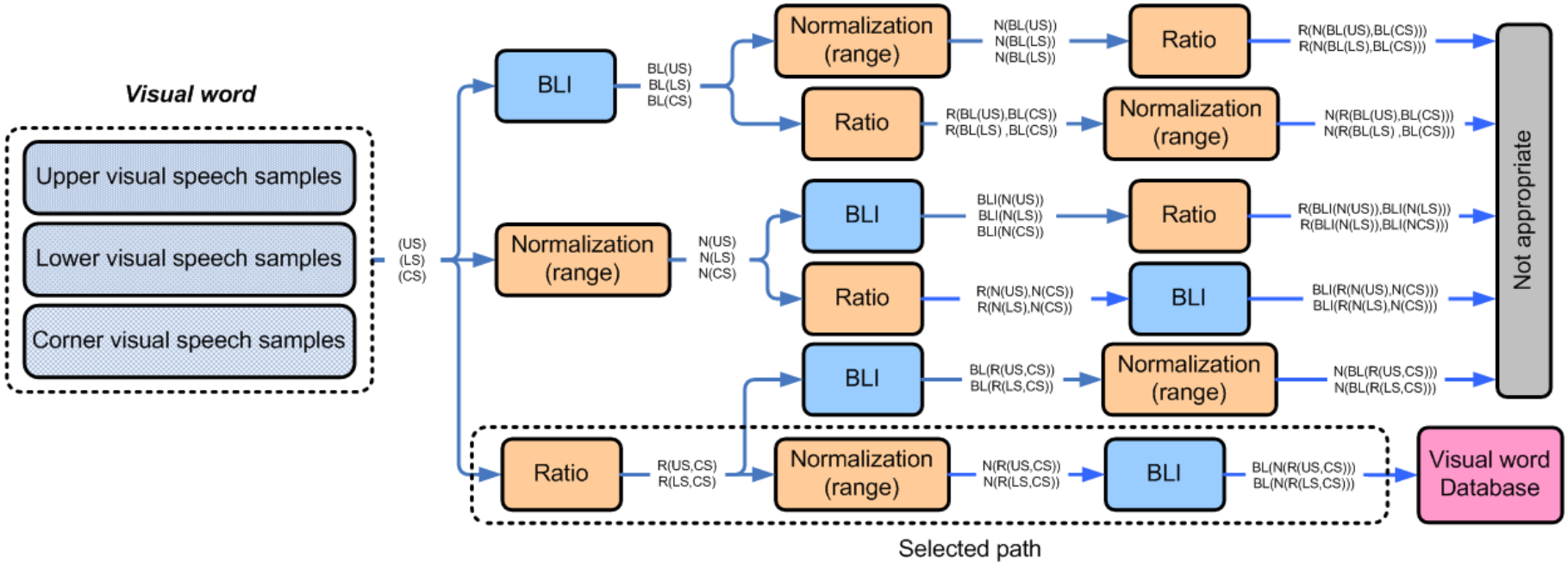


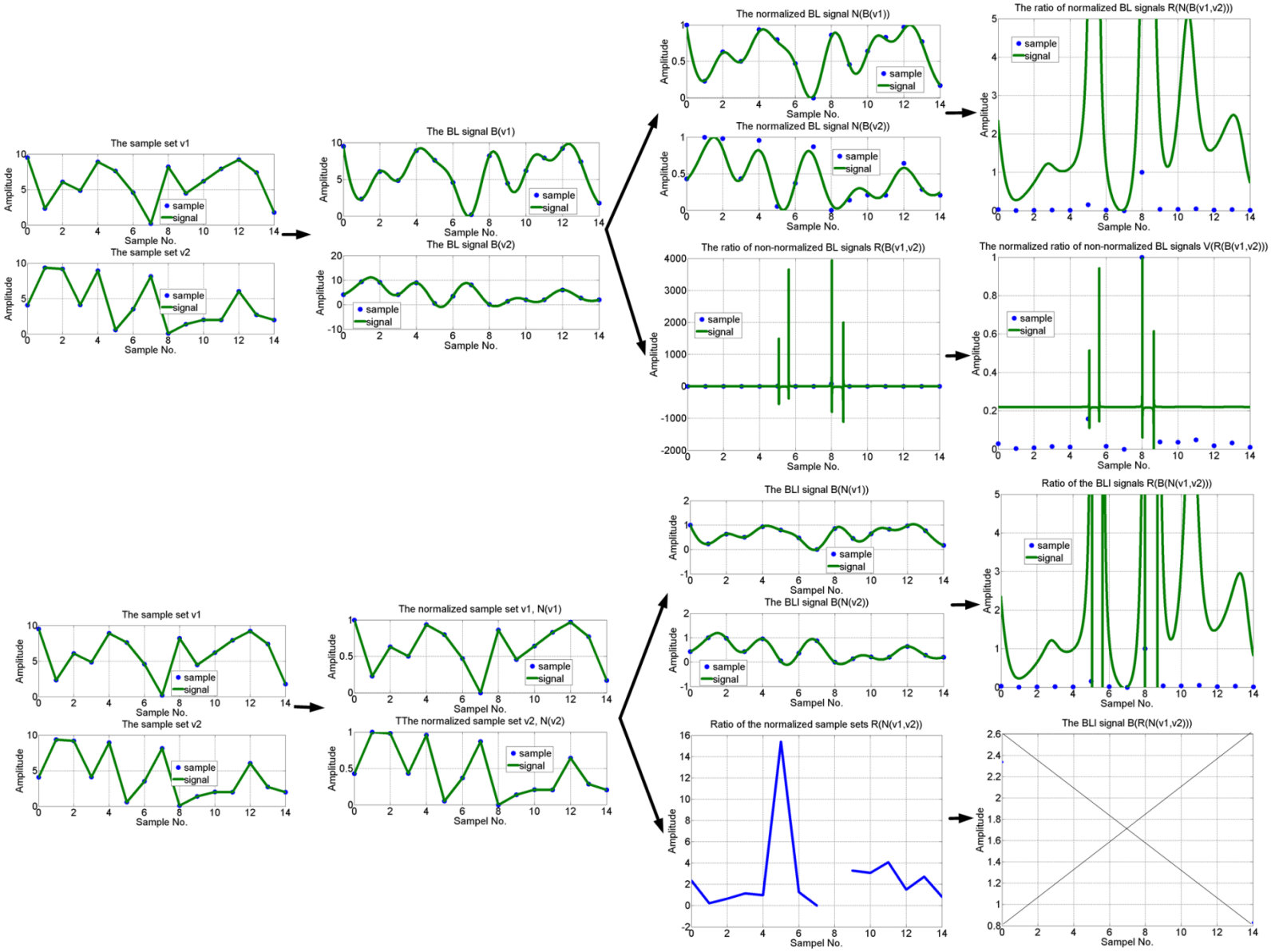
Figure 7-7: The detailed process of mathematical expression generation in all possible configurations

Among all the combination of the BLI, range normalization and rational relating components, there is only one acceptable sequence of them. Deriving the mathematical expressions has been performed using the 'Symbolic Math Toolbox' in the MATLAB environment. For simplicity of referencing the outcomes of each component in order to choose the appropriate combination, the following notations marked the output of each component:

- The Upper, lower and corner visual speech sample sets denoted by US, LS and CS, respectively
- The normalization operation is denoted by$N(.)$
- The ratio of sample sets is denoted by$R(.)$
- The BLI construction is denoted by $BL(.)$

These notations are represented on the connecting line between output and input of each component. For example, the ratio of upper to lower visual speech signals derived from normalized visual speech sample sets is denoted by $R(B(N(US)), B(N(LS)))$.

In Figure 7-7, the graphical representation of all possible combinations are shown. In Figure 7-8 two set consisting of $n = 15$ samples are examined for obtaining the corrct path.. The process of selecting them has been narrowed down by tracking the acceptable results.


The sample set v1
The sample set v2
The BL signal B(v1)
The BL signal B(v2)
The normalized BL signal N(B(v1))
The normalized BL signal N(B(v2))
The ratio of normalized BL signals R(N(B(v1,v2)))
The ratio of non-normalized BL signals R(B(v1,v2))
The normalized ratio of non-normalized BL signals V(R(B(v1,v2)))
The normalized sample set v1, N(v1)
TThe normalized sample set v2, N(v2)
The BLI signal B(N(v1))
The BLI signal B(N(v2))
Ratio of the BLI signals R(B(N(v1,v2)))
Ratio of the normalized sample sets R(N(v1,v2))
The BLI signal B(R(N(v1,v2)))
Amplitude
Sample No.
Sampel No.
sample
signal

(Continues)

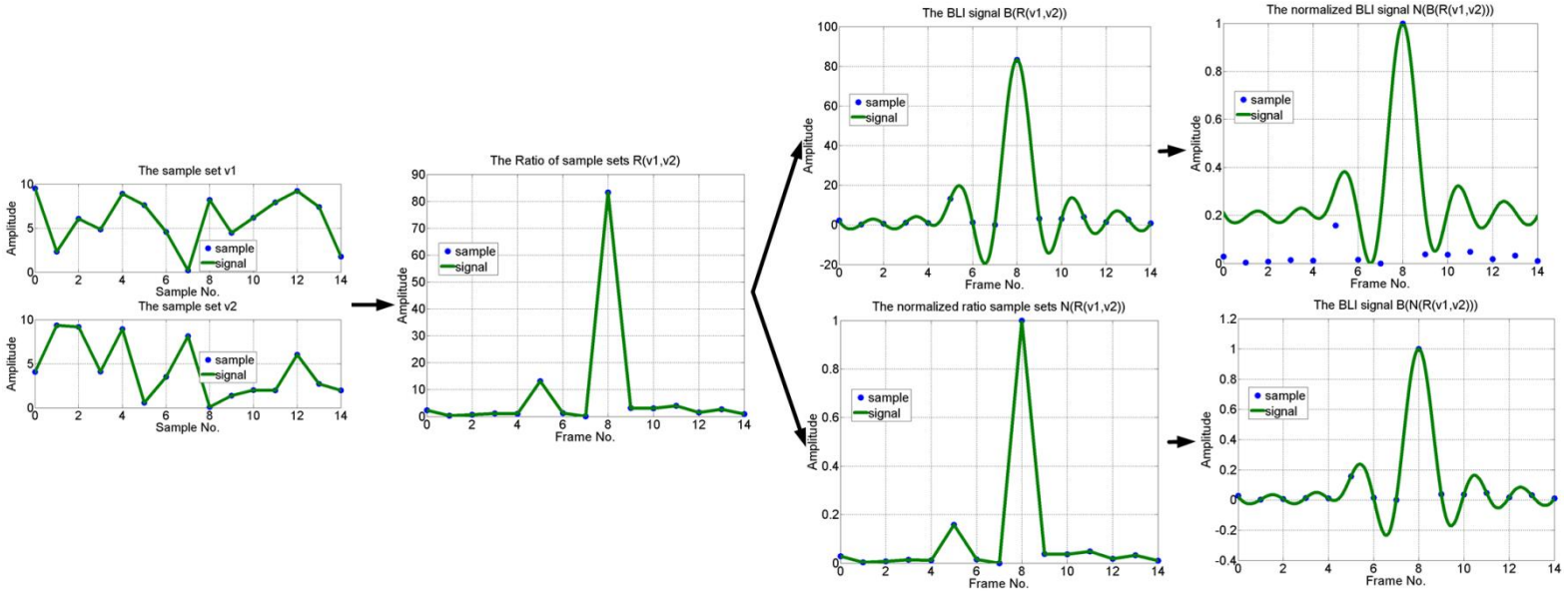


Figure 7-8: The visualized outcomes of all possible configurations as they appear in the output of each stage

From Figure 7-8, the mathematical expressions are derived from the ratio of feature points sequences and then applying to the BLI. It is because finding the expressions from rational function obtained by the ratios of the BL polynomial could result in infinite series expansion.

The next important criterion is the ability of recovering the feature sample sets from the polynomial equations. The combination of normalization and ratio stages has direct effect on the ability of extracting the visual speech sample sets from the formulated visual speech sample sets. On the other hand, the generated expressions should not include any pole into their expressions. Finally, the effect of the BLI converging rate needs to be considered since as it demonstrated in the 'evaluation of the BLI' section there is error between the sample set and interpolating polynomial. These limitations narrow down the choices to a particular combination of components. The particular process of the visual speech signal storage in the mathematical form has been chosen as the sequential combination of visual speech signals rational relations, the normalization and finally BLI, as it shown in Figure 7-9. This combination can be denoted as $BL(N(R(US, CS)))$ and $BL(N(R(LS, CS)))$ to express the visual speech signals relations.

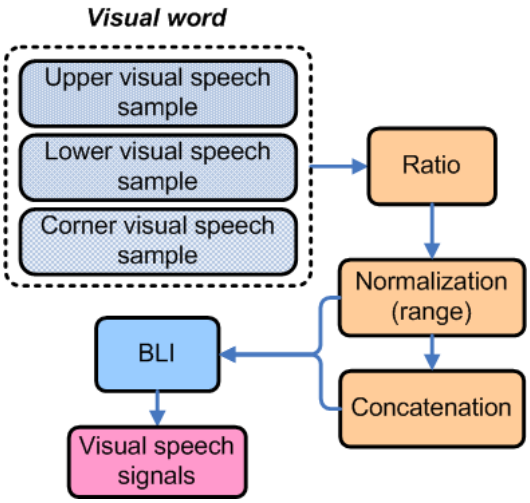


Figure 7-9: The process of visual speech signal generation

Recovering the visual data is possible by extracting the visual speech sample sets from visual speech signals expressions and reversing the process of normalization. Afterward, simply using one of the visual speech sample sets and the recovered ratio, the other visual speech sample sets can be determined. By concatenating the

normalized ratio of visual speech sample sets $N(R(US, CS))$ and $N(R(LS, CS))$ and using the BLI, the unique signature of visual word is obtained. Such definition allocates a single expression to visual words. Recovering the upper and lower visual speech sample set will be possible if the lower visual speech sample sets are stored.

- **VISUAL WORD SIGNATURE EXPRESSIONS**

In this section, the visual speech signals ${VS_{BLI}^{uc}}^{W_m}(f_i)$ and ${VS_{BLI}^{lc}}^{W_m}(f_i), m = \{1,2,3, \dots, 13\}$, which are derived from the normalized ratios of the visual speech sample sets, are subject of visualization. The visual speech signals constructed from the normalized ratios of the upper to corner visual speech sample sets are depicted in Figure 7-10. These visual speech signals have the same trend to the Figure 7-3 to Figure 7-5 since normalization preserves the amplitude trends.

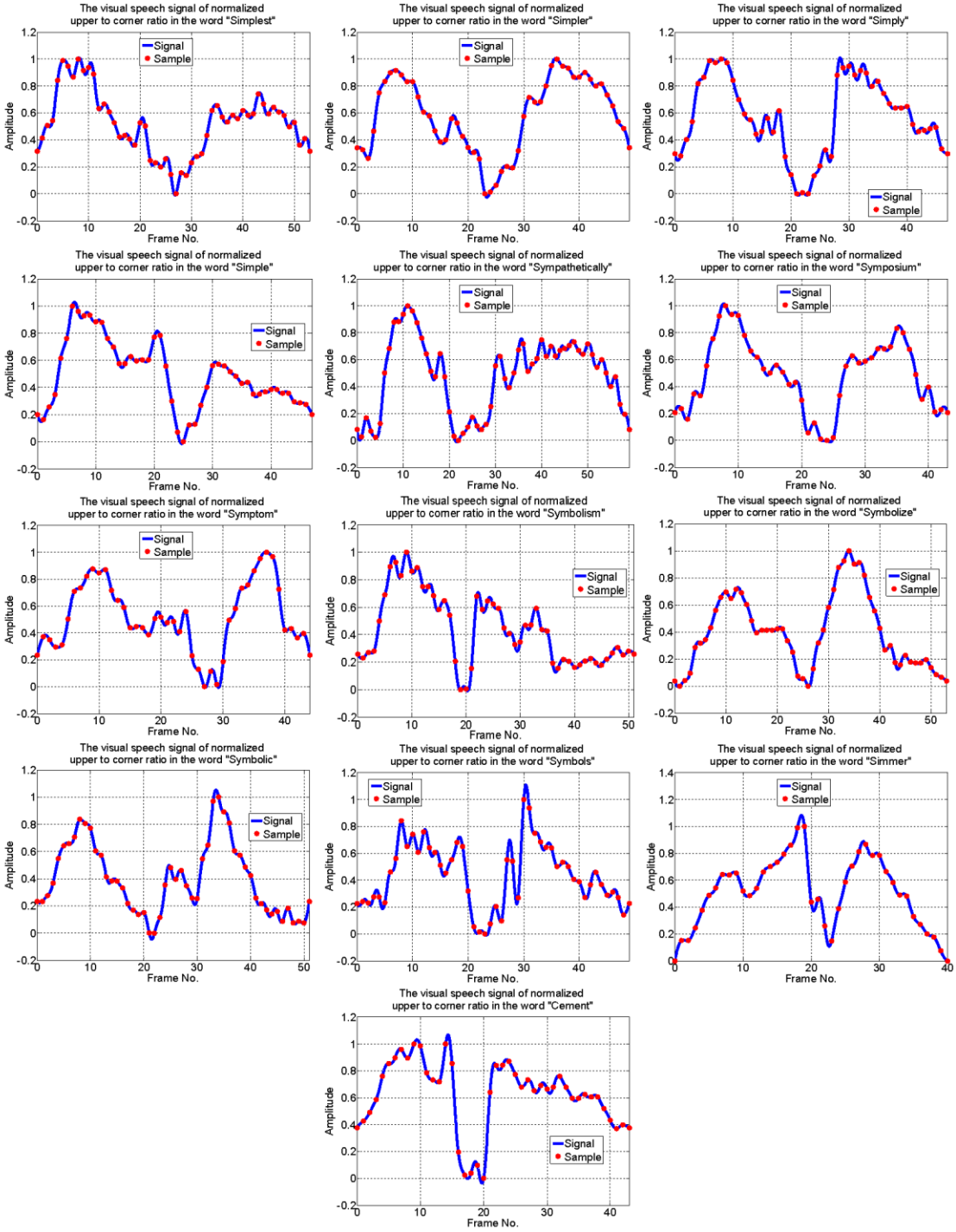


Figure 7-10: The visual speech signals constructed by the normalized ratio of upper to corner visual speech sample sets for all selected words

In Figure 7-11, the process of division, normalization and the BLI is applied to the ratios of the lower to corner visual speech sample sets.

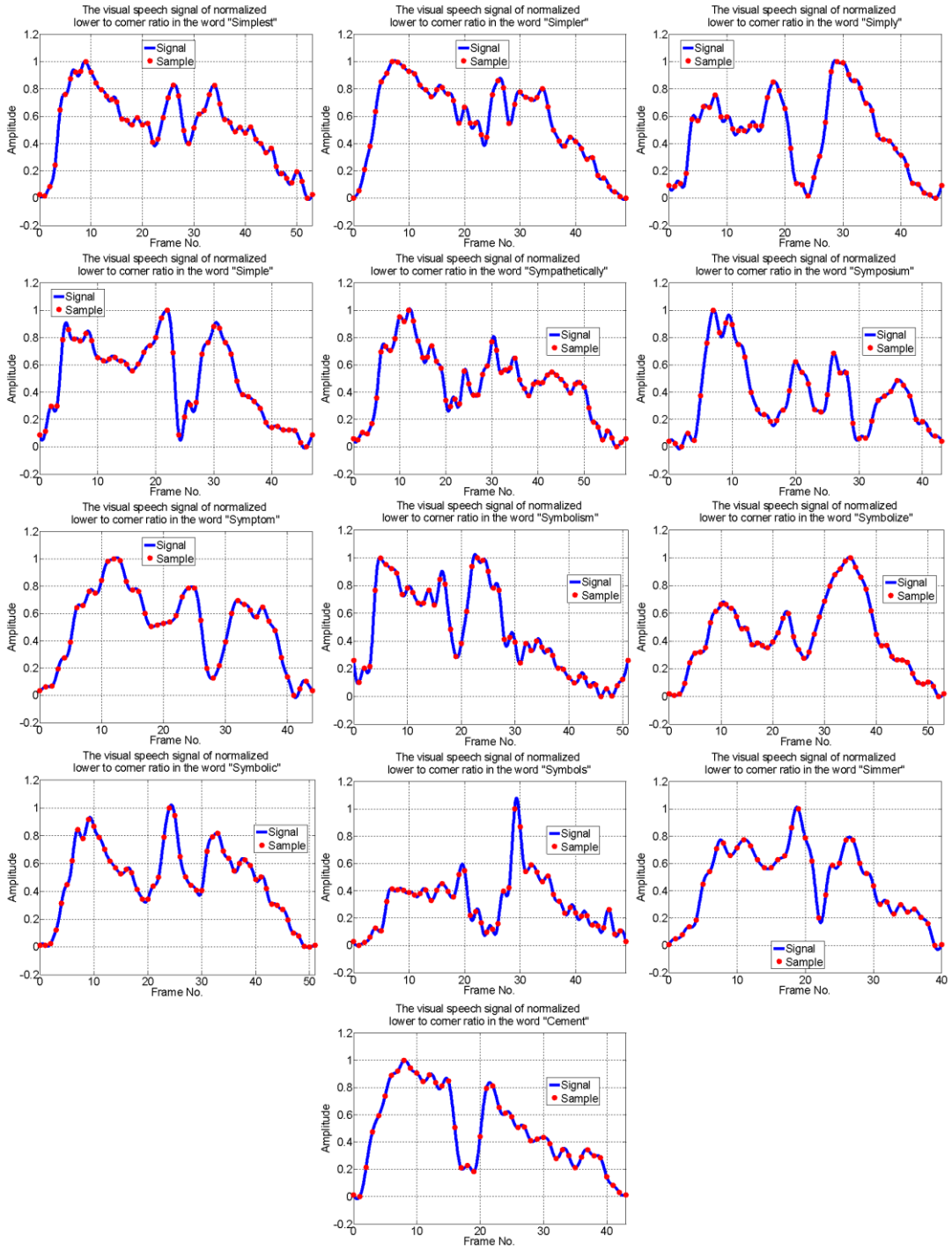


Figure 7-11: The visual speech signals constructed by the normalized ratio of lower to corner visual speech sample sets for all selected words

- **CONCATENATION OF THE VISUAL WORD SIGNATURE EXPRESSIONS**

Concatenating the normalized visual speech sample set and applying the BLI led to a category of the visual speech signal $VS_{BLI}^{uclc^{W_m}}(f_i), m = \{1,2,3,\dots,13\}$ that represents a visual word uniquely but without a mathematical expression due to computational cost complexity. These visual speech signals are shown in Figure 7-12.

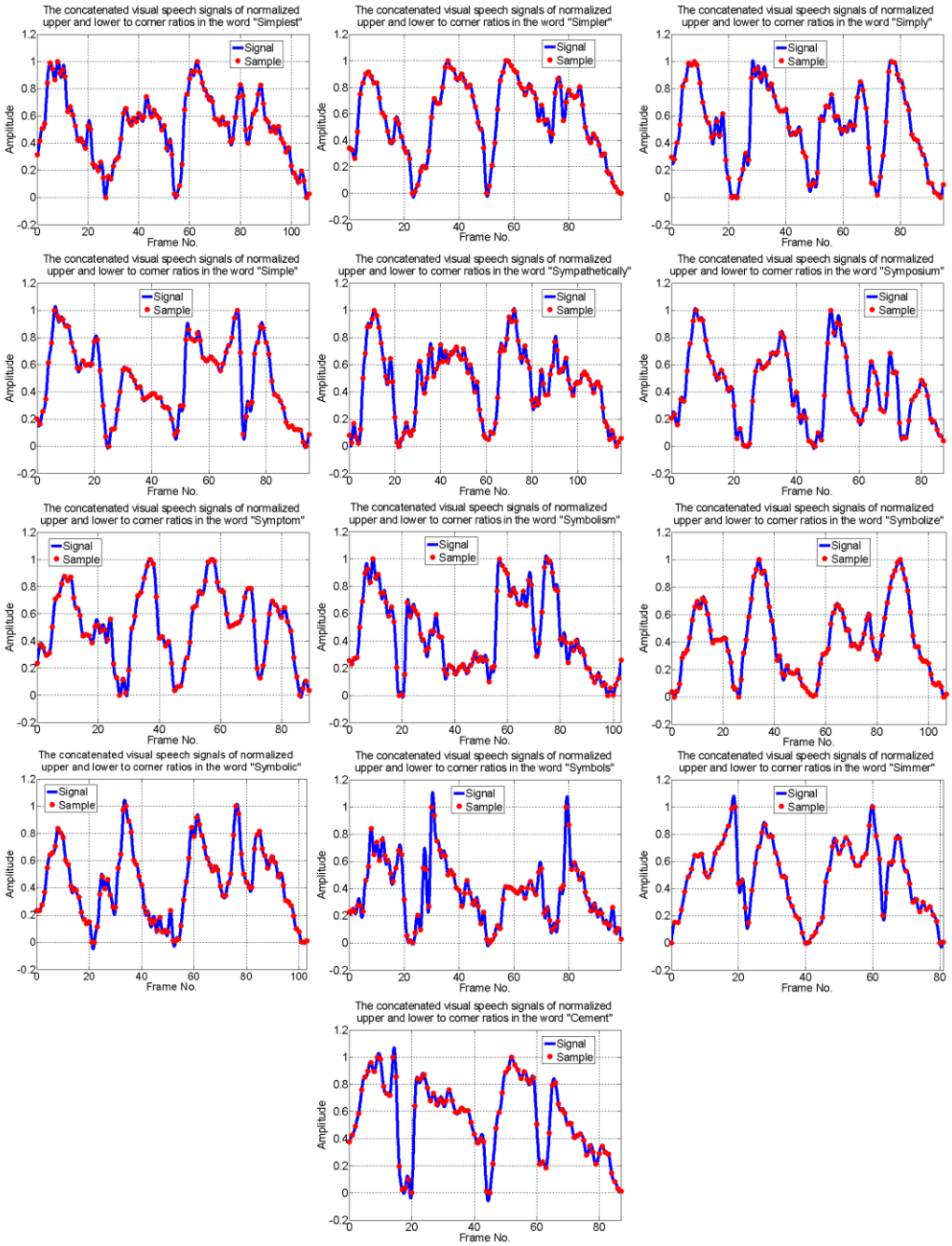


Figure 7-12: The concatenated visual speech signals constructed by the normalized ratio of upper to corner and lower to corner visual speech sample sets for all selected words

7.2. RECOVERING THE VISUAL SPEECH SAMPLE SETS

In this section, the visual speech samples are recovered from the mathematical expressions in order to validate the accuracy of the BLI model. The convergence of interpolating polynomial to the samples is an important issue that needs to be considered for recovering the samples from the mathematical expressions. In this section, the effect of sampling frequency on reconstructing the samples from BL expressions is described. In other words, it is desirable to find a sampling frequency that recovers the lip's movement with a certain error. The BLI convergence concept also depends on the smoothness of the function that provides the samples (Berrut & Trefethen, 2004). Therefore, the convergence needs to be examined according to a particular sample set with priory knowledge. The accuracy of recovering of formulated visual speech signal, e.g. ${VS_{BLI}^{u}}^{W_m}(x)$ completely depends on the sampling frequency over the spans distanced on the frame interval $[0: x_\beta - 1]$. The value x_β corresponds to the number of frames in the visual feature samples of a word.

By substituting the corresponding frame integers $x = x_\beta, x_\beta = \{0, 1, 2, \dots, F_{W_{m_q}} - 1\}$ in ${VS_{BLI}^{u}}^{W_m}(x_\beta)$ written MATLAB code, the outcome is NaN in standard IEEE arithmetic.

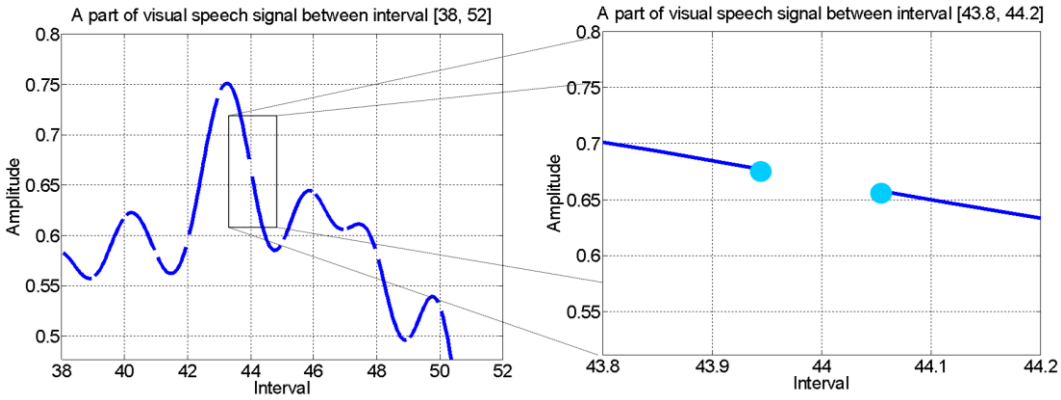
$$ {VS_{BLI}^{u}}^{W_m}(x_\beta) = NaN \tag{7-1} $$

where $x_\beta \in [0: F_{W_m} - 1]$. It is due to the fact of division of the BLI basic functions by zero. Therefore, there is not enough information about the actual amplitude on interval x_β. This issue is addressed by adding a constant decimal fraction ∂_{x_β} to all frames:

$$ x_\beta{}' = x_\beta \pm \partial_{x_\beta} \tag{7-2} $$

The constant ∂_{x_β} ensures recovering of samples with a particular error. Determination of ∂_{x_β} is related to the sampling factor β that is the sampling frequency of visual speech signal (the interpolating polynomial).

For instance, with the sampling frequency $\beta = 20$, the visual speech signal is defined by discrete curve sections as demonstrated in Figure 7-13. By increasing the β factor, the visual speech signal tends to fill the gaps that occurred around the samples amplitudes.



(a) (b)

Figure 7-13: The discontinuity of (a) the BLI signal in the sample interval and (b) detail about a discontinuity in an interval

If $\beta \to \infty$ then $\partial_{x_\beta} \to 0$, but practically it is desirable to find acceptable thresholds, which would preserve the continuity of visual speech signals, as well as accurate recovering of samples.

$$x = \{x_j{}^{\beta}\} \qquad x_j{}^{\beta} = \frac{j}{\beta} \qquad j = 0, 1, 2, \dots, \beta(F_{W_m} - 1) \tag{7-3}$$

At the first glances the idea of extracting the values of the last and first successive portions' samples as the best candidates of actual sample's amplitude would emerge. Therefore, with the error ∂_{x_β} it is possible to choose either value. In this work, as it visualised in Figure 7-14, the suggested method is based on geometrical examination of visual speech signals on interval $x_\beta = i, i \in [1: F_{W_m} - 2]$.

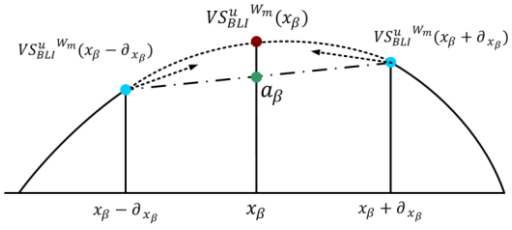


Figure 7-14: The geometrical representation for convergence error determination

$$a_\beta = \frac{(VS_{BLI}^{u}{}^{W_m}(x_\beta - \partial_{x_\beta}) + VS_{BLI}^{u}{}^{W_m}(x_\beta + \partial_{x_\beta}))}{2} \tag{7-4}$$

$$\partial_{x_\beta} = |VS_{BLI}^{u}{}^{W_m}(x_\beta) - a_\beta| \tag{7-5}$$

For the first and the last samples on $x_0 = 0$ and $x_{F_{W_m}-1} = F_{W_m} - 1$, Eq. (7-4) becomes:

$$a_0 = \frac{VS_{BLI}^{u}{}^{W_m}(x_0 + \partial_{x_o})}{2} \tag{7-6}$$

$$a_{F_{W_m}-1} = \frac{VS_{BLI}^{u}{}^{W_m}(x_{F_{W_m}-1} - \partial_{x_{F_{W_m}-1}})}{2} \tag{7-7}$$

The maximum error for each word is chosen as the index of BLI accuracy for reconstruction of sample set:

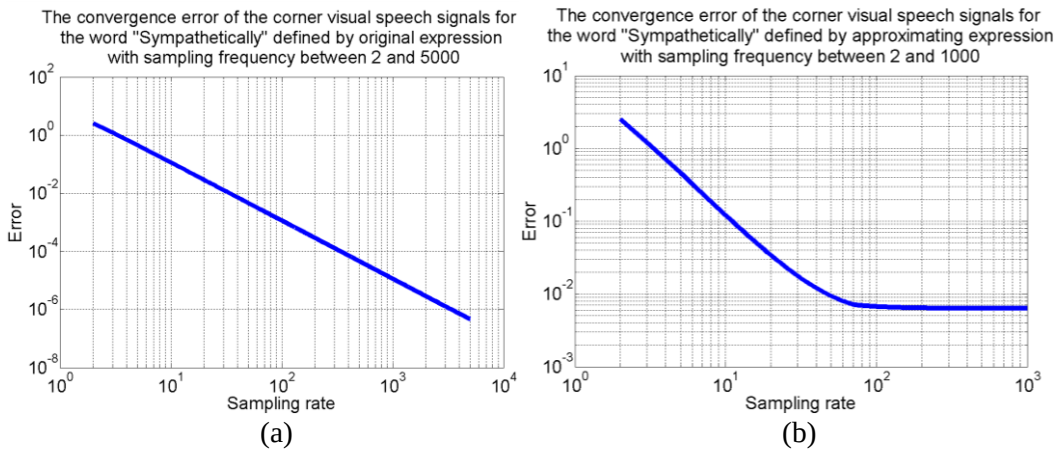
$$\partial_{W_m} = max\{\partial_{W_m}\} \tag{7-8}$$

$$\partial_{W_m} = max\{\partial_{W_m}\} \tag{7-9}$$

The geometrical determination of amplitudes provides the maximum error in each mathematical expression. When the sampling frequency in an expression increases, the error ∂_{W_m} reaches to a smaller value and cause more accurate sample recovery. In order to regenerate the sample sets from expressions, the decimal fraction ∂_{W_m} is adding to $x_\beta = i$ to obtain:

$$x_\beta{}' = i + \partial_{W_m} \tag{7-10}$$

for $i \in [0, F_{W_m} - 1]$. In this work, the mathematical expressions of the visual signals are represented with decimal fraction of four digits. The reason of such approximation is due to the extremely extended mathematical expression. The result of such approximation reveals the mathematical expressions reach to a region where increasing the sample rate does not affect error reduction.



(a) (b)

Figure 7-15: The comparison between the convergence error rate of word 'sympathetically' for the (a) original corner visual speech signal with sampling rage between 2 and 5000 and (b) the approximated corner visual speech signal with sampling rage between 2 and 1000

For instance, in Figure 7-15 (a) the mathematical expression of the corner visual speech signal appeared in the word 'sympathetically' before approximating with decimation of four digits is represented from $\beta = 2$ to 5000 sampling. The error is constantly reduces to reach its minimum in approximately 4.7×10^{-7}, while in Figure 7-15 (b) the corner visual speech signal decimated of four digits reaches to a constant error approximately equal to 6.4×10^{-3}.

- **RECOVERING THE VISUAL SPEECH SAMPLE SETS**

The convergence error of the mathematical expressions of the upper, lower and corner visual speech signal as well as the visual speech signal derived from the normalized ratios of the upper to corner and lower to corner visual speech sample sets are represented in this section. The convergence error has been calculated from the maximum difference between the geometrical approximation of the samples and the actual visual speech sample sets and their normalized ratios versus increasing the sampling rate to 1000 samples in the mathematical expressions (see Figure 7-16 to Figure 7-20).

- **CONVERGENCE ERROR IN THE UPPER VISUAL SPEECH SIGNALS** $VS_{BLI}^{u}{}^{W_m}(f_x)$

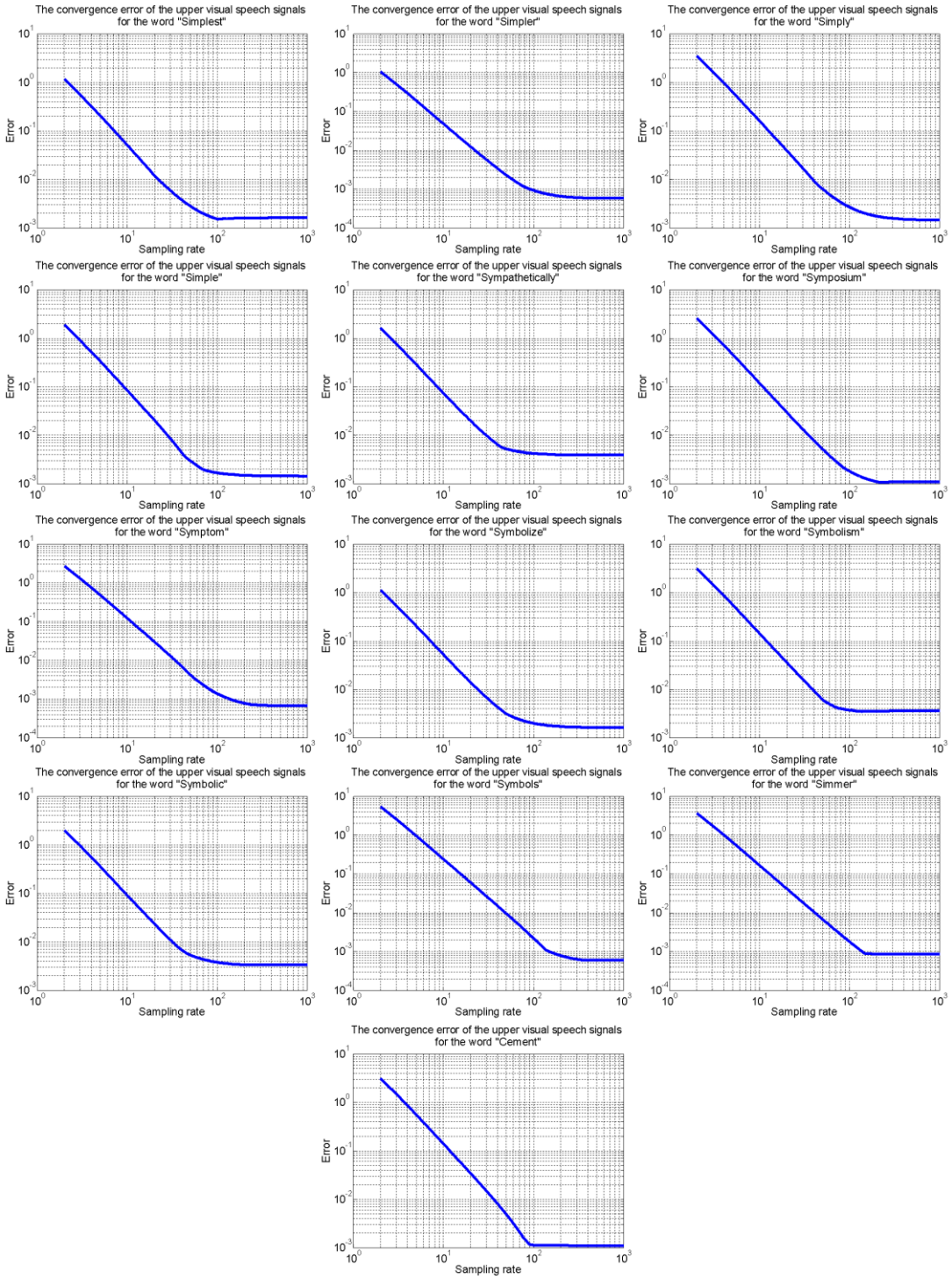


Figure 7-16: The BLI convergence errors in reconstruction of the upper visual speech signals

- **CONVERGENCE ERROR IN THE LOWER VISUAL SPEECH SIGNALS** $VS_{BLI}^{l}{}^{W_m}(f_x)$

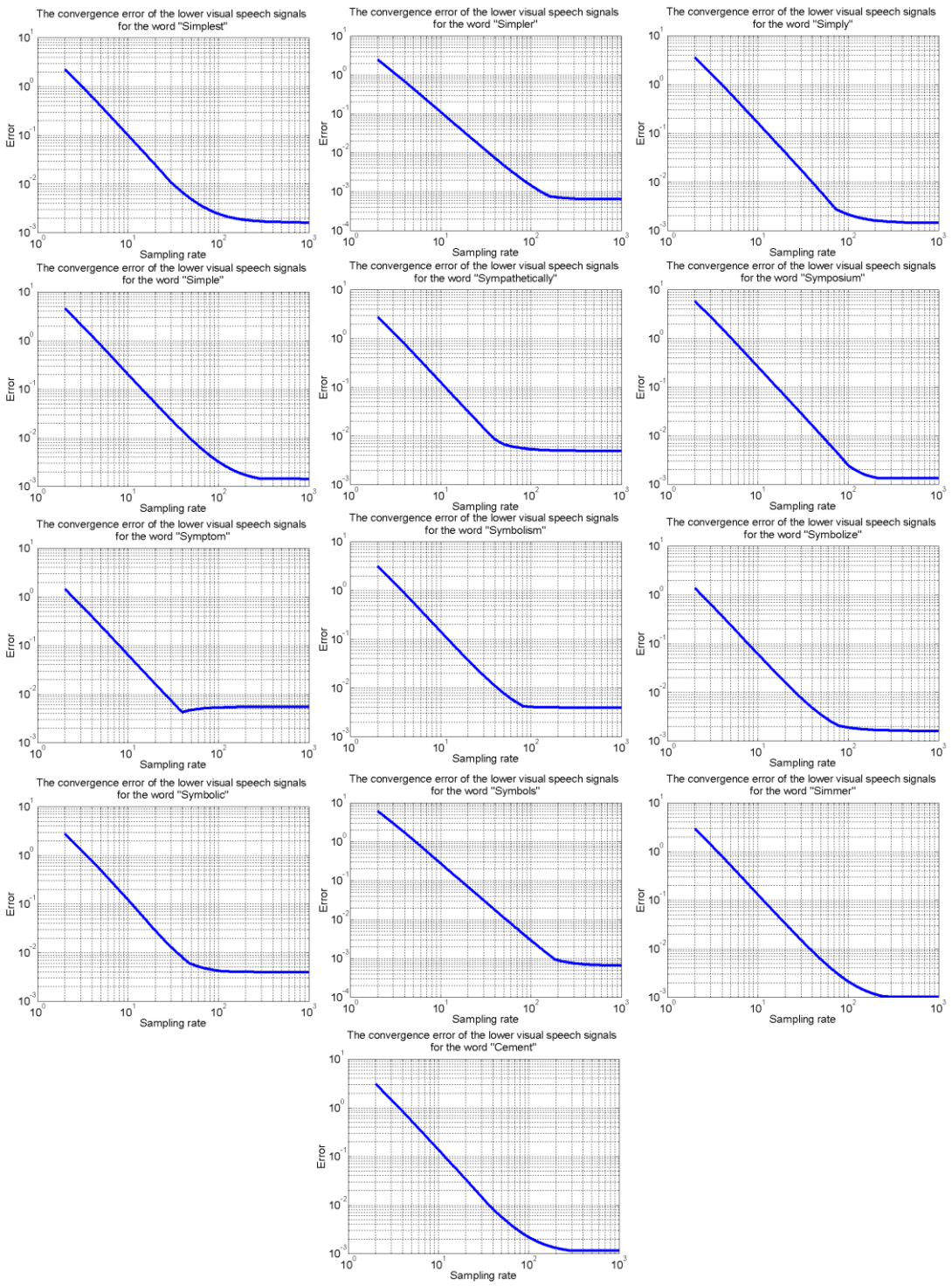


Figure 7-17: The BLI convergence errors in reconstruction of the lower visual speech signals

- **CONVERGENCE ERROR IN THE CORNER VISUAL SPEECH SIGNALS** $VS_{BLI}^{c}{}^{W_m}(f_x)$

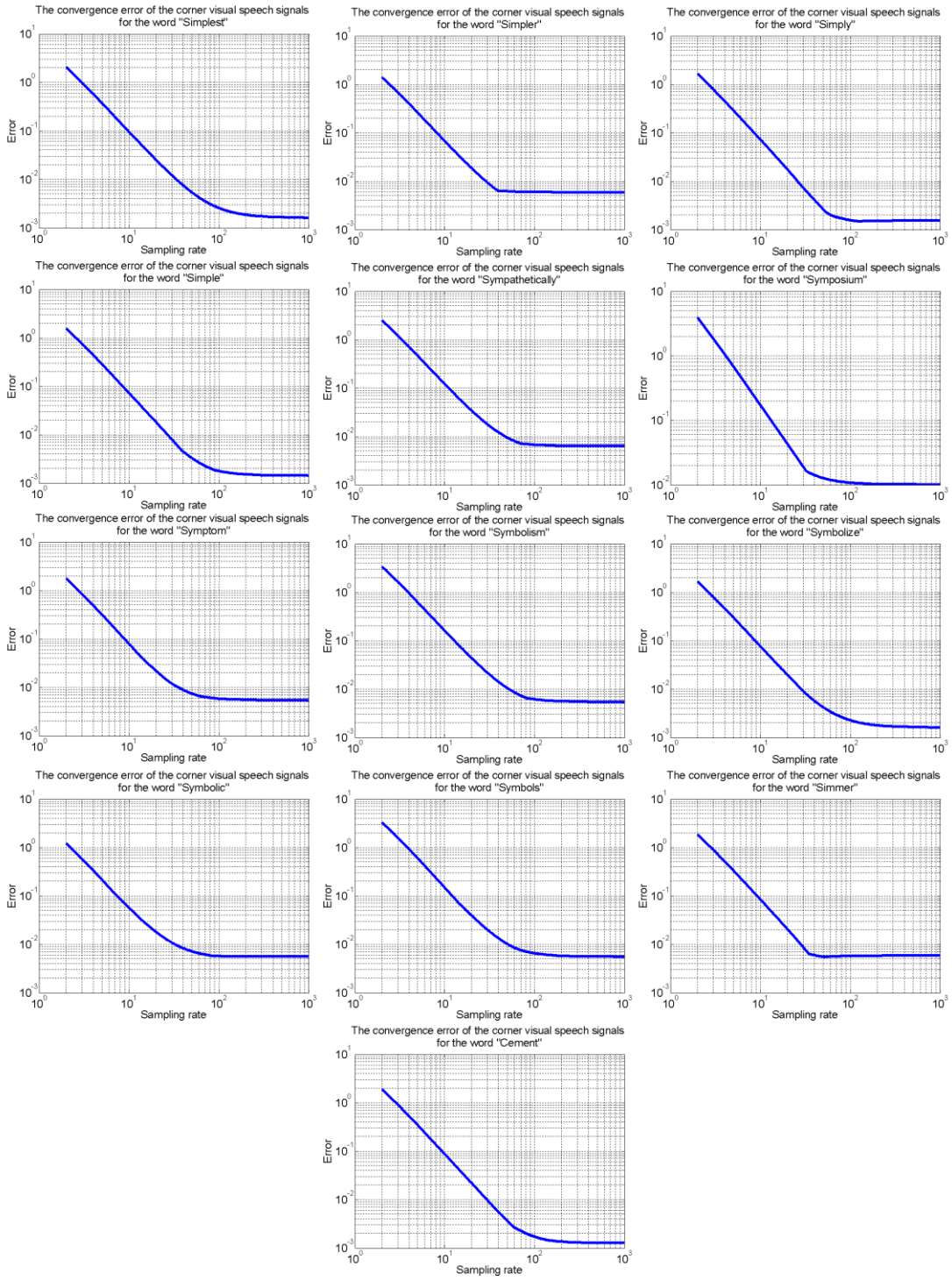


Figure 7-18: The BLI convergence errors in reconstruction of the corner visual speech signals

- **CONVERGENCE ERROR IN THE VISUAL SPEECH SIGNALS** $VS_{BLI}^{uc}{}^{W_m}(f_x)$

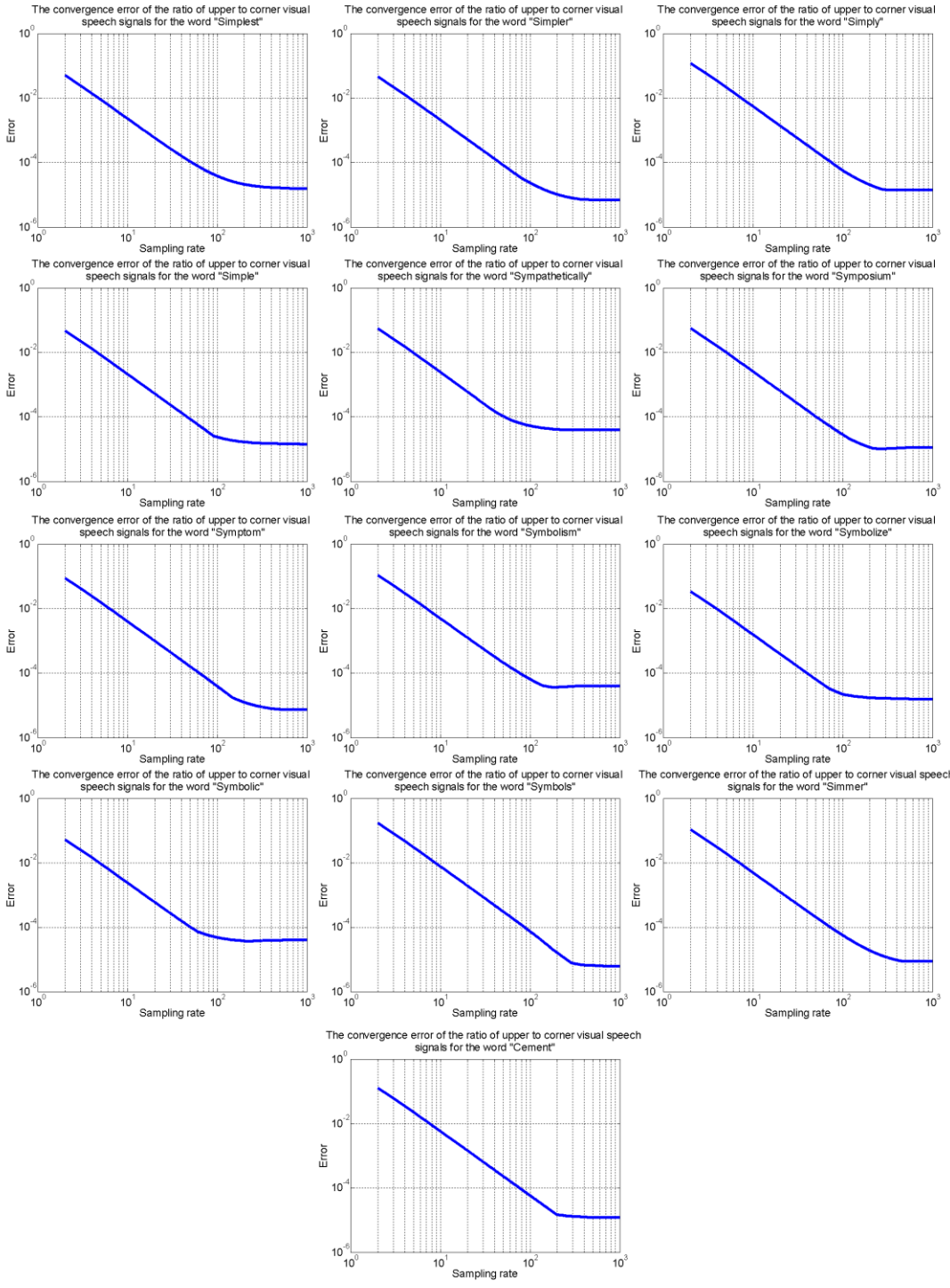


Figure 7-19: The BLI convergence errors in reconstruction of the visual speech signals obtained from the normalized ratio of upper to corner visual speech sample sets

- **CONVERGENCE ERROR IN THE VISUAL SPEECH SIGNALS** $VS_{BLI}^{lc\ W_m}(f_x)$

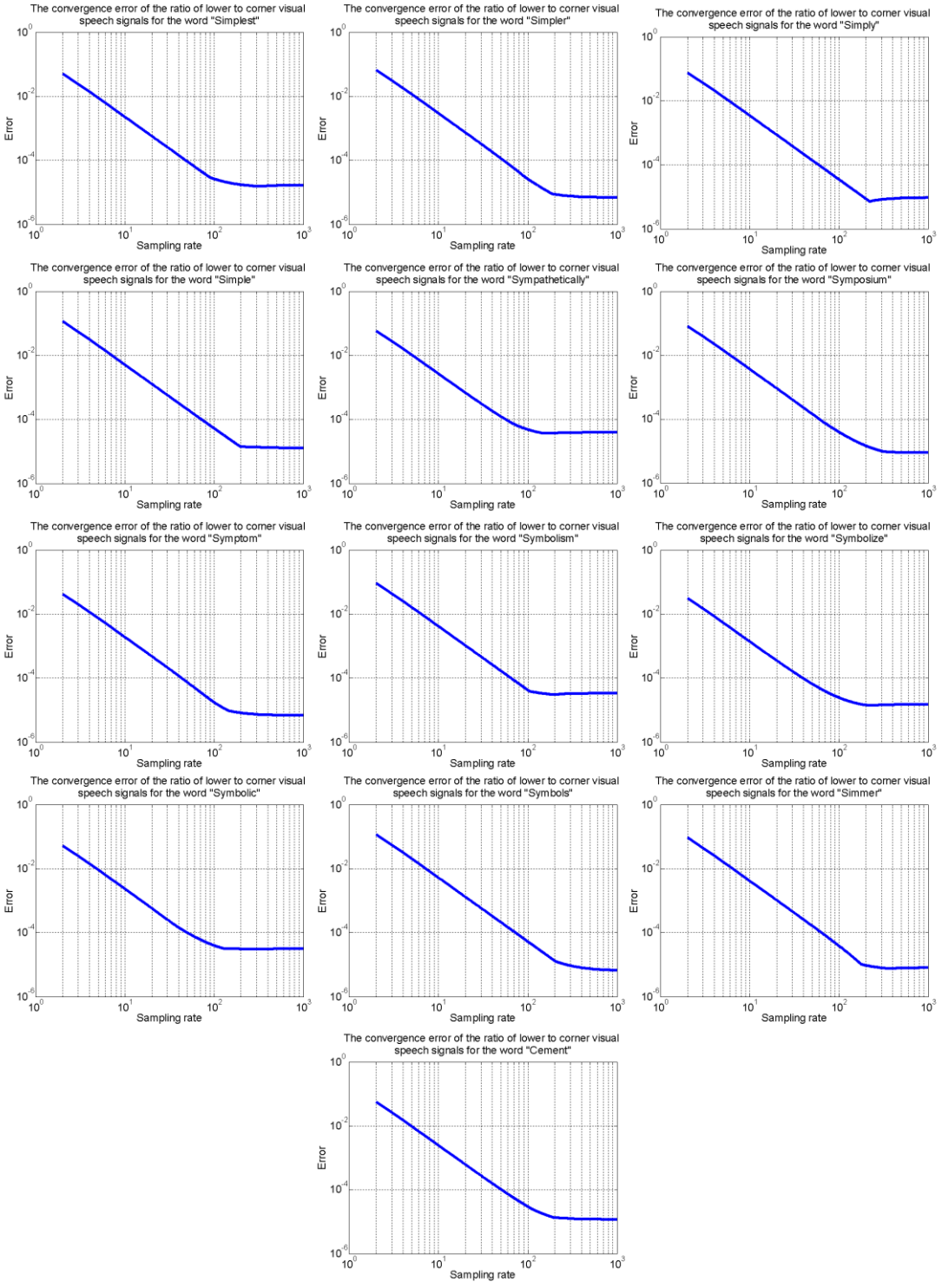


Figure 7-20: The BLI convergence errors in reconstruction of the visual speech signals obtained from the normalized ratio of lower to corner visual speech sample sets

By increasing the sample rate near 200, the recovery error reduces while further increasing the sample rate do not improve the recovering ability of sample sets. However, the error in the lower visual speech sample set for the 'Symptom' (Figure 7-17) increases after reaching to a minimum with 40 samples.

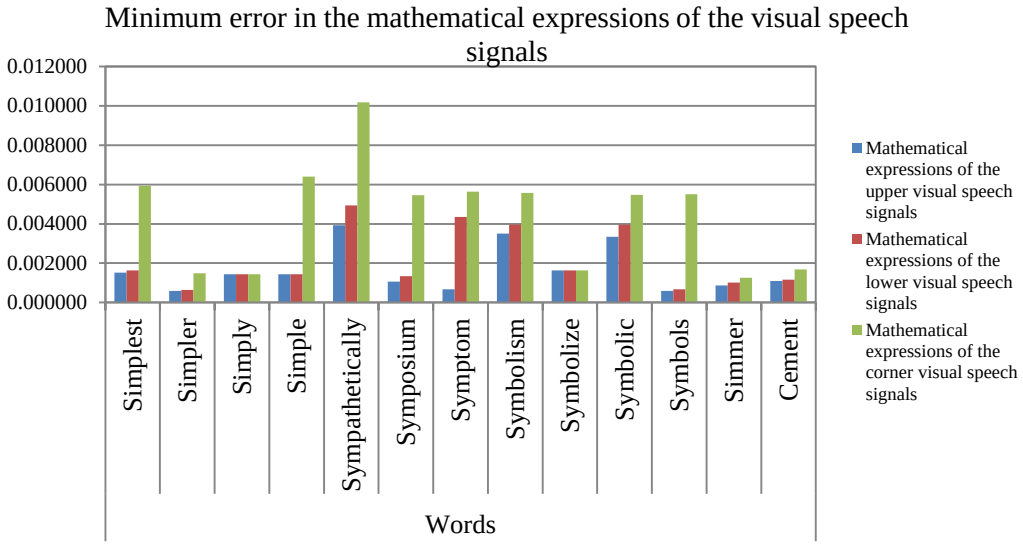


Figure 7-21: The minimum errors for the mathematical expressions of the visual speech signals $\boldsymbol{VS_{BLI}^{u}}^{W_m}(\boldsymbol{f_x})$, $\boldsymbol{VS_{BLI}^{l}}^{W_m}(\boldsymbol{x})$ and $\boldsymbol{VS_{BLI}^{c}}^{W_m}(\boldsymbol{f_x})$

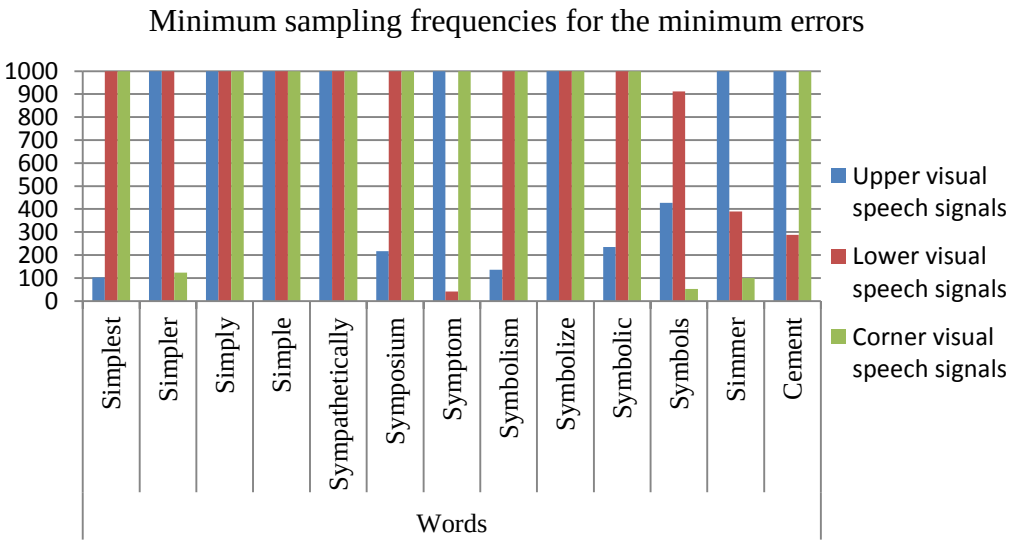


Figure 7-22: The minimum sampling frequencies for the minimum errors

The error and the corresponding sampling rates are shown in Figure 7-21 and Figure 7-22. The minimum error for recovering the upper visual speech sample set in Figure

7-21 is 5.84×10^{-4} while the maximum error is 39.33×10^{-4}. Similarly, for the lower visual speech data the minimum error is 6.47×10^{-4} and maximum error is 49.32×10^{-4}. The recovered corner visual speech has minimum and maximum error equal to 12.66×10^{-4} and 101.8×10^{-4}, respectively.

Table 7-1: The minimum and maximum error and corresponding sampling rates in recovered visual speech sample sets

			Error	Sampling rate	Words
Recovered visual speech sample sets	Upper	Min	5.84×10^{-4}	103	Simpler
		Max	39.33×10^{-4}	1000	Sympathetically
	Lower	Min	6.47×10^{-4}	41	Simpler
		Max	49.32×10^{-4}	1000	Sympathetically
	Corner	Min	12.66×10^{-4}	52	Simmer
		Max	101.8×10^{-4}	1000	Sympathetically

On the other hand, the corresponding minimum and maximum sampling rate for the recovered upper, lower and corner visual speech sample sets are 103, 1000, 41, 1000, 52 and 1000 samples. The minimum and maximum error and corresponding sampling rates in recovered visual speech sample sets are represented in Table 7-1.

7.3. BARCODE ALLOCATION

For the reason of standardization and identification, a specific platform, which allocates a set of more simplified rules to the visual speech signals, needs to be selected. One of the potential identification methods is allocating barcode to each signature polynomial. According to the international standard of verification barcode verification defined as: "The technical process by which a barcode measured to determine its conformance with the specification for that symbol that deals with symbology specification and also any additional specification (Silva & Oney, 2009)". The aim of verification is to check ability to fulfil the symbol to its function (Silva & Oney, 2009). Barcodes are arrangements of bars with variable edge, length, and spacing that represent a set of encoded symbols. The length and spacing are the subject of encoding and decoding of symbols. In this section, the suggested method allocates a barcode to the signature polynomial by employing the *pulse width*

modulation (PWM). The PWM, which uses amplitude comparison between the signal and carrier, is a method of modulation. The carrier signal is chosen as sequence of triangles (triangle waveform). The modulated signal is a sequence of rectangular waveforms with constant amplitude and different durations.

The period of triangle pulse train is matched to the BL polynomial period. It leads to defining a triangle pulse train where each triangular pulse's central frequency is located on frame i covering $\frac{f_i - f_{i+1}}{2}$, $i = 0, 1, 2, \dots, N-1$ on minimum amplitudes. Since the samples are spanned uniformly, the width W in Eq. (7-11) is equal to one.

The peak-to-peak amplitude of triangle pulses tri_{P-P} is scaled to cover the range of the BL polynomials denoted by BL_{P-P} that result in a modified version of triangular pulse train tri_s using:

$$tri_s(x) = BL(x) \times tri(x) + min[BL(x)] \tag{7-11}$$

where $BL(x)$ is the BL polynomial. The result of comparison between $BL(x)$ and $tri_s(x)$ is a PWM signal denoted by $BL^{PWM}(x)$ detected by the following expression.

$$BL^{PWM}(x) = \begin{cases} Max[BL(x)] & if \quad BL(x) \le tri_s(x) \\ min[BL(x)] & if \quad BL(x) > tri_s(x) \end{cases} \tag{7-12}$$

The sampling frequency also should be equal to the sampling frequency of the BL interpolating polynomial otherwise the comparison between the BL polynomials and triangle pulse train could not be possible. As it shown in Figure 7-23, a signal is compared with a triangular pulse denoted with dashes. The bar's length is defined based on the distance between coinciding points (denoted by circles). The length of each bar τ_i is called 'on' time that varies between $[0, 1]$ on the length of each span T_i, so:

$$0 \le \tau_i \le 1 \tag{7-13}$$

for $i = \{0,1,2, \dots, (N-1)\}$.

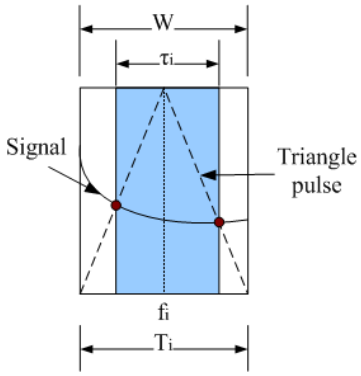


Figure 7-23: The geometrical representation of barcode generation from a part of signal with aid of a triangle pulse (dashed)

The on time τ_i is restricted between the signal and triangle pulse coinciding points that is marked by red circles. The minimum and maximum points of the signal $BL(x)$ are denoted by $Max[BL(x)]$ and $min[BL(x)]$ where defined the minimum and maximum for the triangle pulse train. The $BL(x)$ polynomial is covered with the $tri_s(x)$ waveform denoted with circles. By replacing the BL polynomial with the desirable visual speech signals, the associated barcode not only simplifies them but also can be processed for allocating identification properties for the visual speech signals.

The block diagram in Figure 7-24 visualized the process of barcode generation. For relating the visual speech signals, the ratio block is used along with construction and range normalization process. The PWM block generates the modulated visual speech signal and after normalization of its domain, the barcode would be generated. Before finalizing the barcode generation process, the effect of range normalization would be useful.

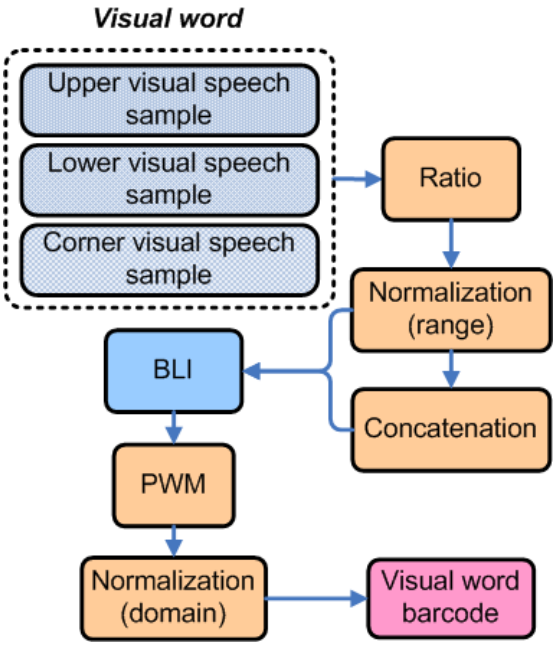


Figure 7-24: The block diagram for visual speech barcode allocation

In Figure 7-25 (a), a constructed signal by the BLI from $n = 10$ random numbers $n \in [0: 10]$ taken from uniform distribution and the triangle waveform are generating the modulated signal that is compared with the constructed signal in Figure 7-25 (b). Similarly, the normalized versions of the signal and triangle waveform Figure 7-25 (c) are defined the modulated signal in Figure 7-25 (d).

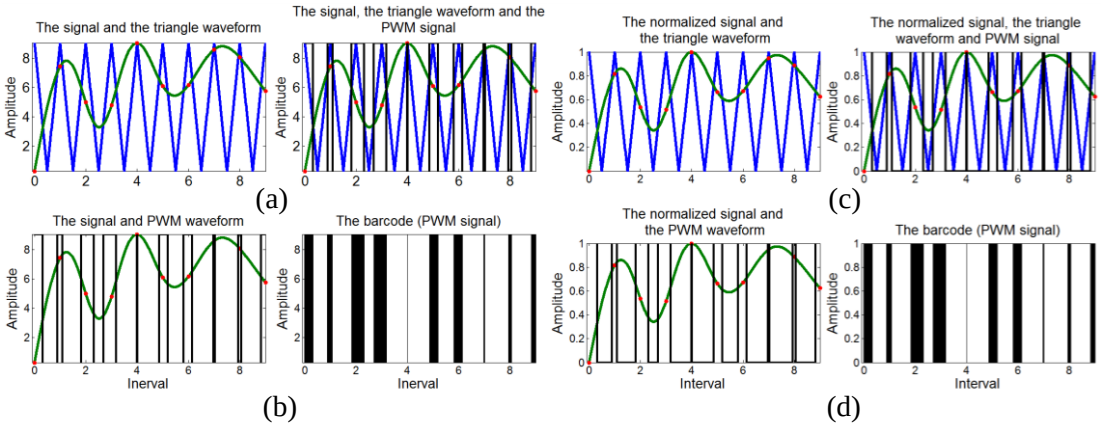


Figure 7-25: The effect of (a) non-normalized and (b) corresponding barcode and (c) normalized signals (d) corresponding barcode

In the next section, in order to simplify the visual signals the barcode properties will be examined.

- **BARCODE PROPERTIES**

For selecting the suitable interaction between the visual speech signals and their barcode representation, it is important to study the properties that could be used as an index of identifications. Like the process of deriving the signature expression (relating the feature points to each other), a very basic property in identification of object is the simplicity of identification rules or relations. Defining more simplified identification rules for barcodes would increase the identification and recovering accuracy. For instance, if the barcode is represented with a number, the memory reservation would be increased. Using these simplified identities of the visual speech signals and related barcodes as an interface, a simplified relation could be obtained. In other words, by calling the properties it does not need to use demodulation techniques to generate the visual speech signals. Obviously, the concept of optimal relation changes according under different circumstances and only after testing the complete design of system it could be resolved. However, in this stage, two basic properties of barcodes, which are simple enough to be considered as 'barcodes identities' are discussed. These identities are derived from the properties on domain of barcodes. It has been observed that the barcode inherits two important properties that can be used to differentiate the objects as the number of bars and also the duty cycle as it demonstrated in Figure 7-26.

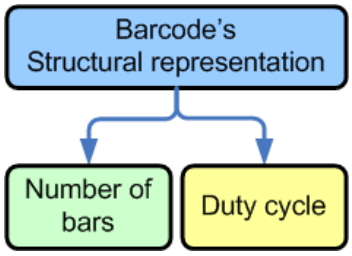


Figure 7-26: The barcode structural properties

These properties are using for further simplifications on the barcode representation. In the following sections, the suggested simplification methods have been described in more details.

- **NUMBER OF BARS**

One of the immediate observations of barcode is the variations of the number of bars with the value of one. This characteristic could be reliable when the length of

extracted samples from lip is not the same. Based on the number of bars appearing in the barcodes, the visual speech sample sets intervals of seven words according to table 6 are categorized in three groups with same lengths. In the first group, the words 'simplest' and 'simply' have 54 samples, in the second group the words 'symbolism', 'symbols' and 'symptom' consisted of 47 samples and in the third group the words 'simple' and 'simpler' consisted of 48 samples. The remaining six words have different number of samples. Therefore, the words barcodes could be categorized in nine groups. Now, it is possible to assume the barcode is prepared and the task is to determine the possibility of identifying the correct group of barcodes. Since 6 words out of 13 words are uniquely, determine the number of words bars, the probability of correct identification is $\frac{6}{13}$. However, there is 53.84% possibility of identify non-unique group of words. In other words, the possibility of correct identification is reliable on 46% of the total number of trials that is not a desirable amount for adopting the number of bars as one of the identification properties.

It is also possible to make the system to be sensitive to those three specific groups. These numbers then would be considered as values that cause identification uncertainty.

- **BARCODE'S DUTY CYCLE**

The other more reliable characteristic of generated barcode is obtained by considering the time's ratio that the barcode remains high to the period of barcode. Such ratio is referred to as barcode's duty cycle (BDC) in this work. The barcode period T is obtained by the number of frames N by the length of spans T_i as:

$$T = N\,T_i \tag{7-14}$$

Afterward, the barcode's duty BDC cycle simply is obtained by dividing the length of each bar τ_i by the barcode period T as:

$$BDC = \frac{\tau_i}{T} \qquad 0 < BDC < 1 \tag{7-15}$$

The possibility of observing equal BDCs is extremely low since all of the polynomials signatures have different amplitudes and consequently they have different times τ_i. The factor of the BL polynomials amplitudes are playing a very critical role on the

nature of generated barcodes. According to that, barcodes can be generated via normalized or non-normalized BL polynomials. In this situation a question about the different generate barcodes related to the similarity between them comes to mind. The normalization of amplitudes as described could be thought as scaling their range between zero and one. But how could it affect the timing of bars? The answer would be provided shortly by examining a simple example and using the BDC.

In this section, the graphical results of barcode allocation to the visual speech signals generated by the normalized ratio of the upper to corner and lower to corner visual speech sample sets as well as their concatenated versions are represented. In Figure 7-27, $VS_{BLI}^{uc}{}^{W_m}(f_i), m = \{1, 2, 3, \dots, 13\}$ the triangular waveform $tri_s(f_i)$ and the resulted PWM signals are depicted, respectively. In Figure 7-28, the barcodes defined by the PWM signals from the visual speech $VS_{BLI}^{uc}{}^{W_m}(f_x)$ are represented.

- **VISUAL SPEECH SIGNALS BARCODE** $BVS_{BLI}^{uc}{}^{W_m}(f_i)$

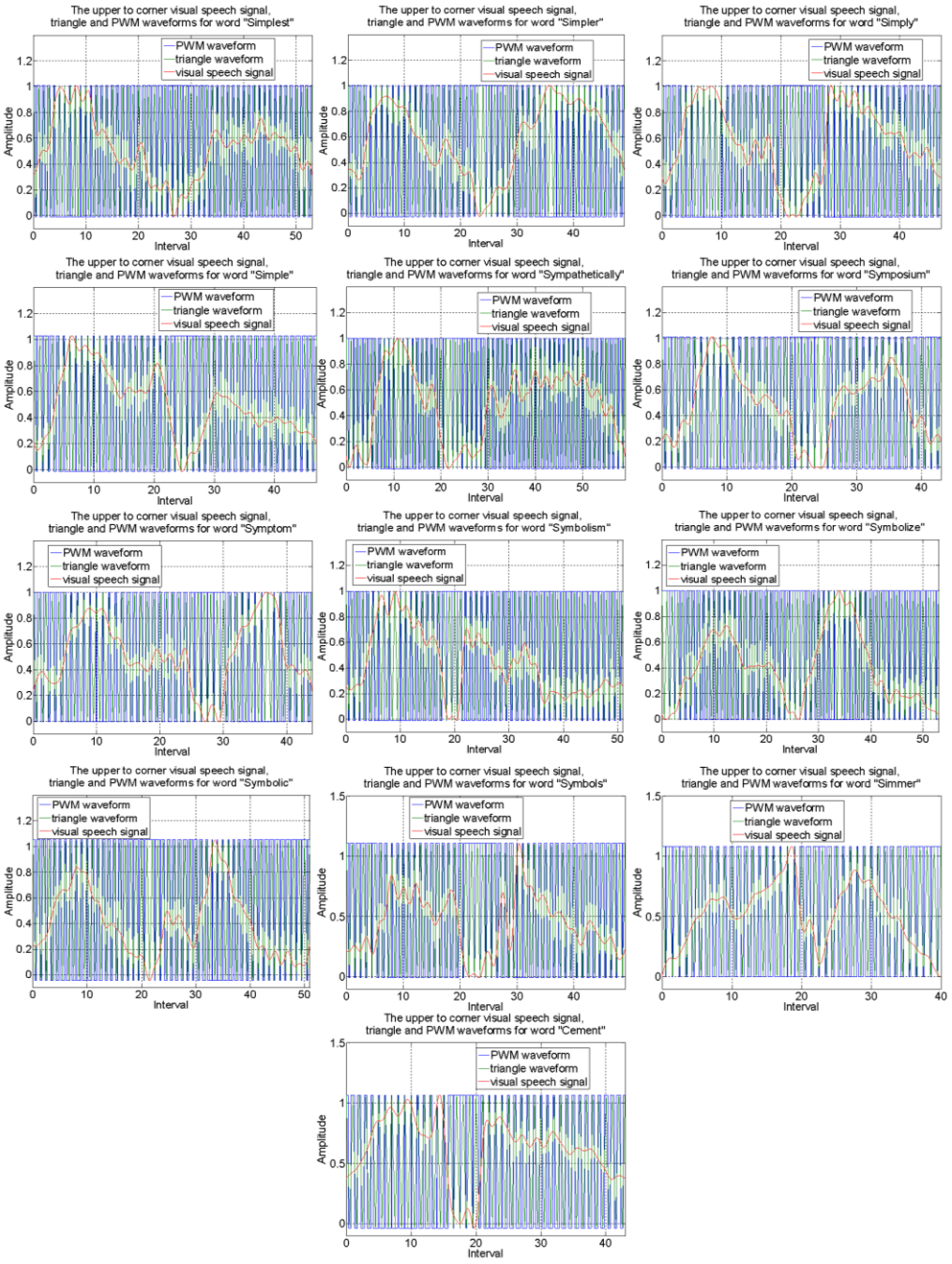


Figure 7-27: The visual speech signals constructed by the normalized ratio of upper to corner visual speech sample sets, the carrier and the PWM waveforms for all selected words

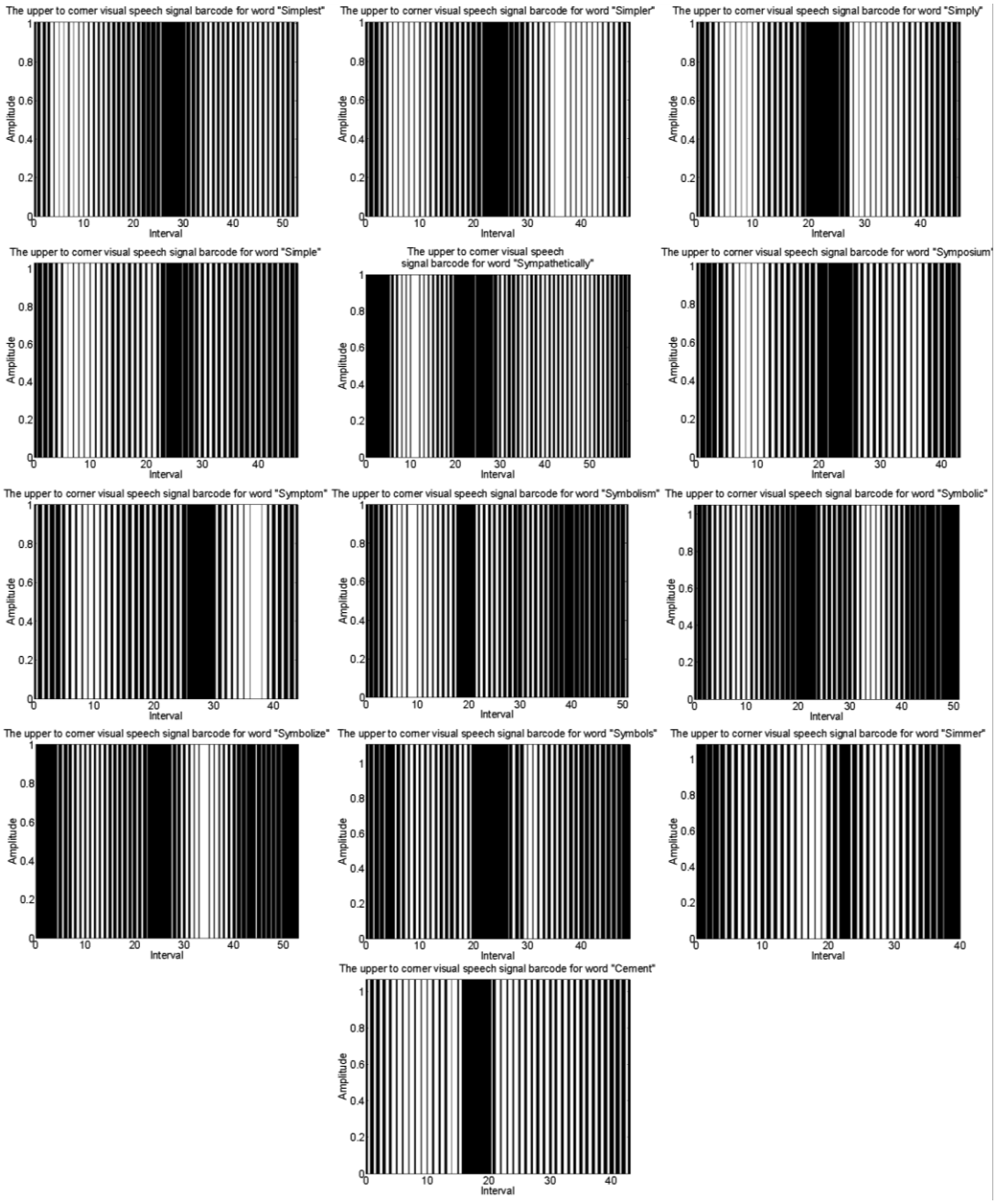


Figure 7-28: The barcode of the visual speech signals constructed by the normalized ratio of upper to corner visual speech for all selected words

In Table 7-2, the barcodes duty cycles of the visual speech signal constructed from the normalized ratio of upper to corner visual speech sample sets denoted by NRUCBDC (Normalized Ratio of Upper to Corner Barcodes Duty Cycle) are calculated for characterizing the visual words by a unique value.

Table 7-2: The barcode duty cycles of the visual speech signals constructed from the normalized ratio of upper to corner visual speech sample sets for all selected words

NRUCBDC	
Simplest	0.474815
Simpler	0.406933
Simply	0.42875
Simple	0.510556
Sympathetically	0.509778
Symposium	0.508182
Symptom	0.476
Symbolism	0.558718
Symbolize	0.597654
Symbolic	0.592949
Symbols	0.594533
Simmer	0.522439
Cement	0.399545

- **VISUAL SPEECH SIGNALS BARCODE** $BVS_{BLI}^{lc}{}^{W_m}(f_i)$

in Figure 7-29, the barcodes defined by the PWM signals from the visual speech $VS_{BLI}^{uc}{}^{W_m}(f_i)$ are presented while in Figure 7-30, $VS_{BLI}^{lc}{}^{W_m}(f_i), m = \{1, 2, 3, \dots, 13\}$ the triangular waveform $tri_s(x)$ and the resulted PWM signals are depicted.

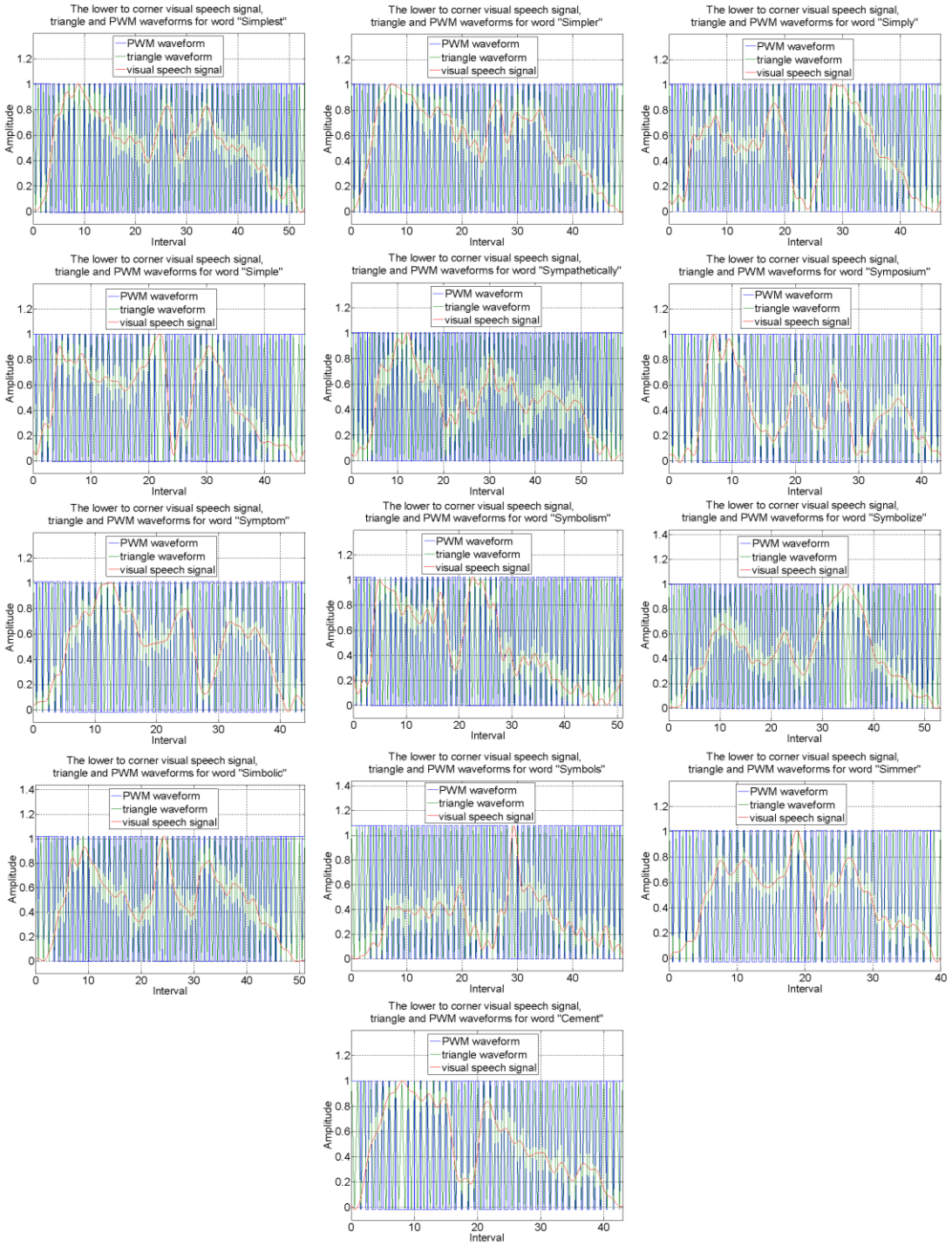


Figure 7-29: The visual speech signals constructed by the normalized ratio of lower to corner visual speech sample sets, the carrier, and the PWM waveforms for all selected words

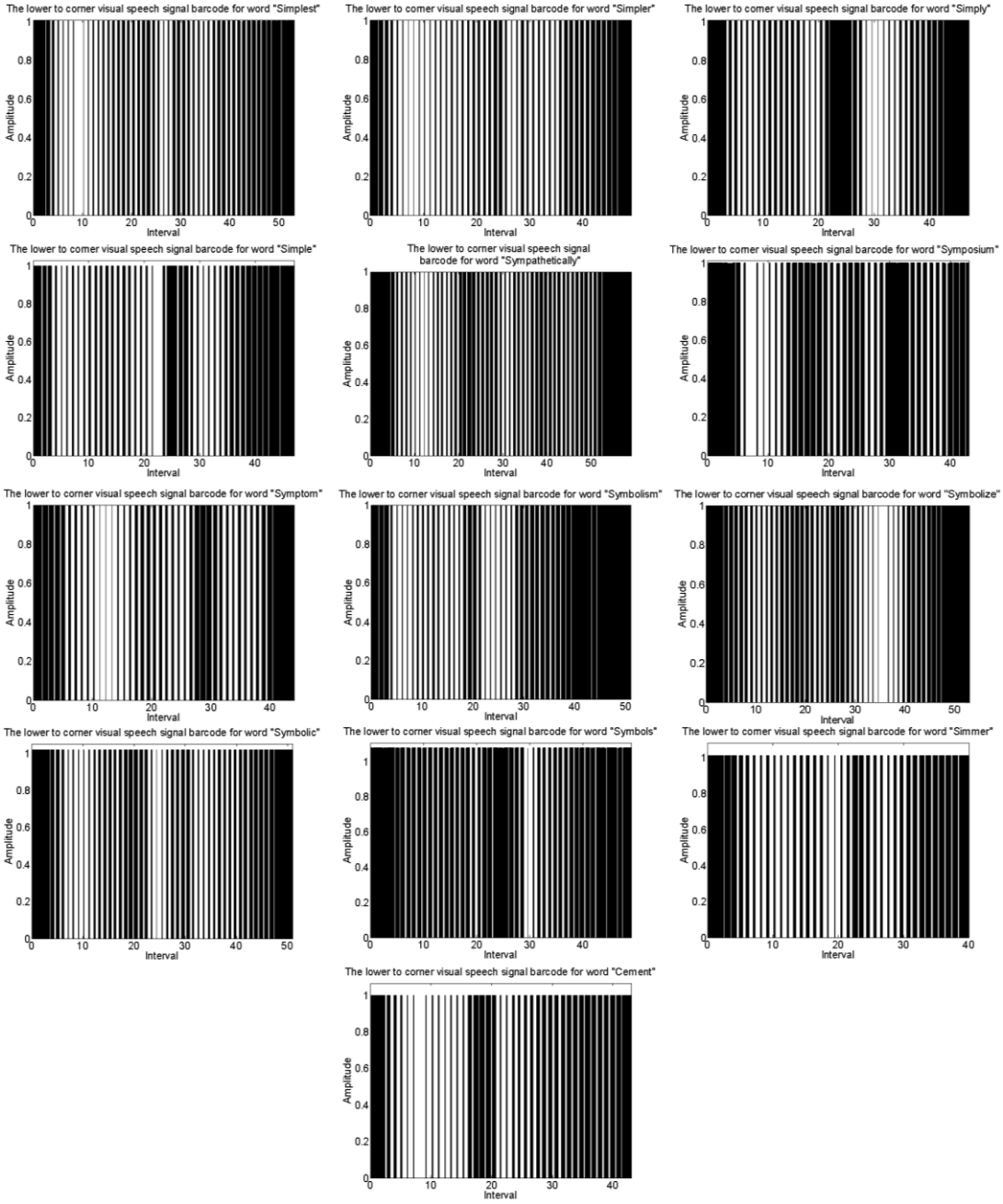


Figure 7-30: The barcode of the visual speech signals constructed by the normalized ratio of lower to corner visual speech for all selected words

In Table 7-3, the barcodes duty cycles (NRLCBDC) are calculated for characterizing the visual words by a unique value.

Table 7-3: The barcode duty cycles of the visual speech signals constructed by the normalized ratio of lower to corner visual speech for all selected words

NRLCBDC	
Simplest	0.47037
Simpler	0.426667
Simply	0.537222
Simple	0.491944
Sympathetically	0.540444
Symposium	0.614091
Symptom	0.48563
Symbolism	0.54141
Symbolize	0.563827
Symbolic	0.51641
Symbols	0.704267
Simmer	0.516098
Cement	0.498939

- **VISUAL SPEECH SIGNALS BARCODE** $\boldsymbol{BVS_{BLI}^{uclc^{W_m}}(f_i)}$

The barcode concept is also allocated to the concatenated visual speech signals that are representing the visual words by a numerical signal. In Figure 7-31, these visual speech signals, the triangular waveform $tri_s(f_i)$ and the resulted PWM signals are illustrated. In Figure 7-32, the final representation of the visual words barcodes are visualized where the corresponding duty cycles ULCBDC (Upper Lower Concatenated Barcodes Duty Cycle) are represented in Table 7-4.

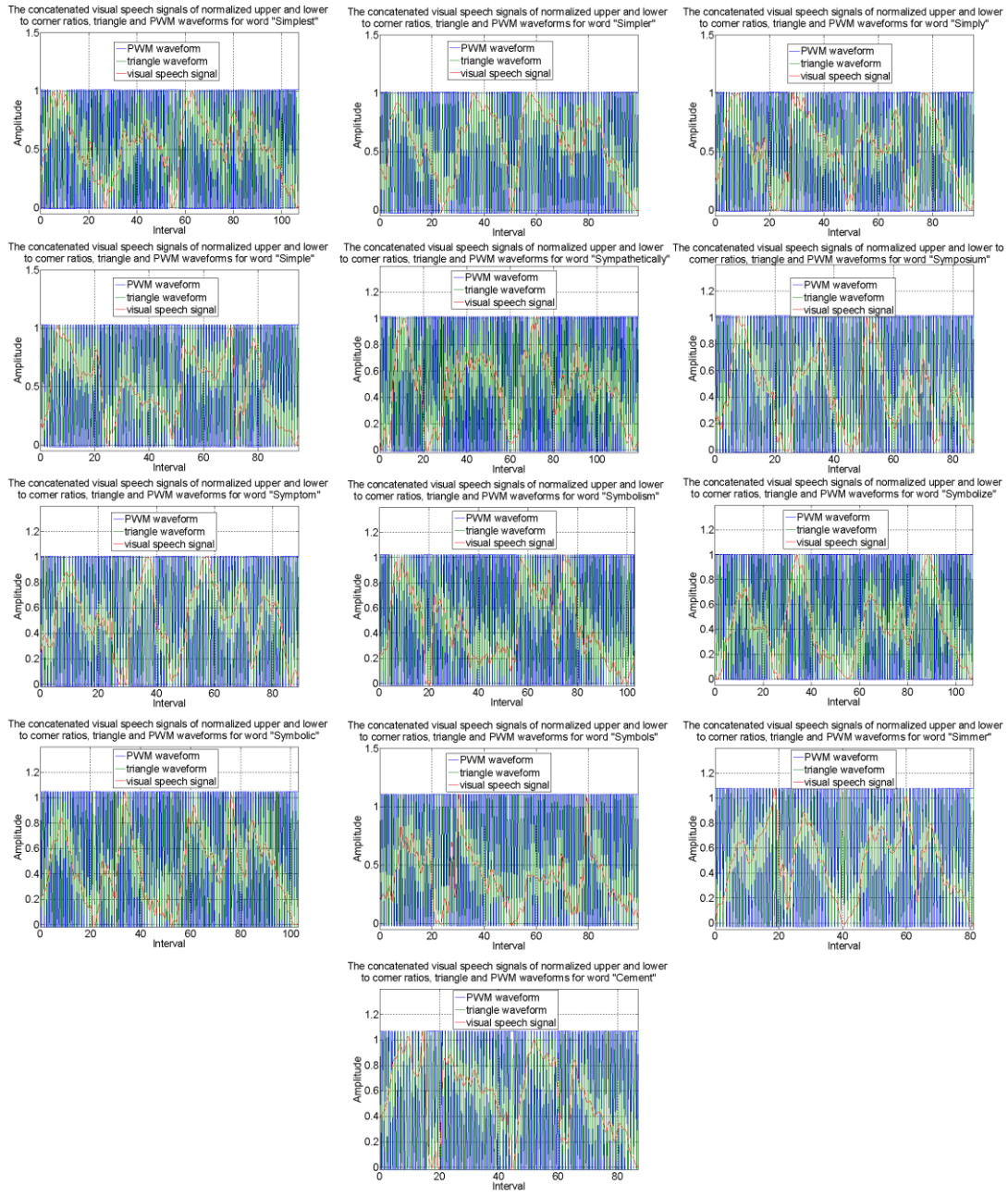


Figure 7-31: The concatenated visual speech signals constructed by the normalized ratio of upper to corner and lower to corner visual speech sample, the carrier, and the PWM waveforms for all selected words

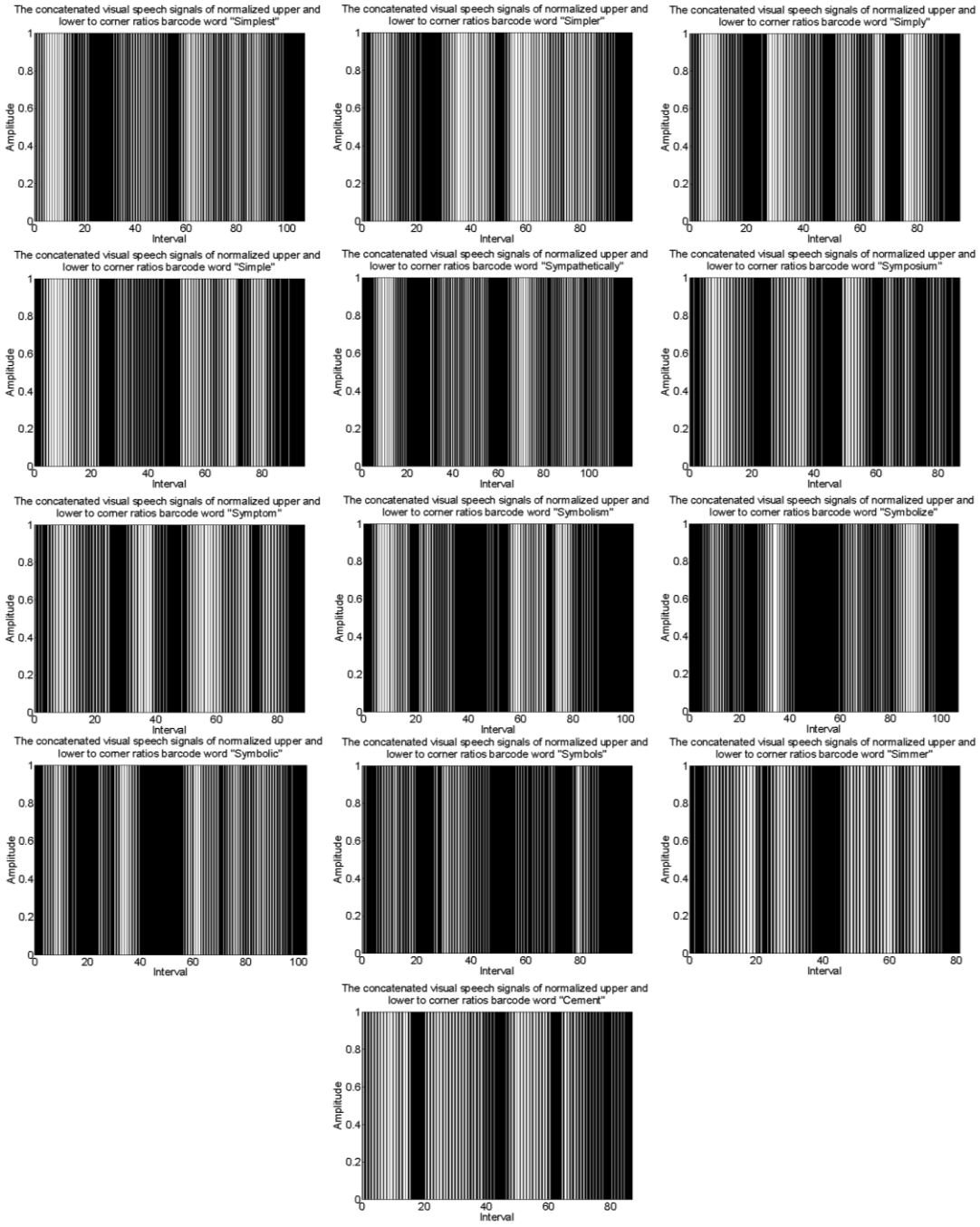


Figure 7-32: The barcodes of the concatenated visual speech signals constructed by the normalized ratio of upper to corner and lower to corner visual speech sample for all selected words

Table 7-4: The barcode duty cycles of the concatenated visual speech signals constructed by the normalized ratio of upper to corner and lower to corner visual speech sample for all selected words

ULCBDC	
Simplest	0.478704
Simpler	0.418467
Simply	0.483472
Simple	0.508472
Sympathetically	0.530222
Symposium	0.565833
Symptom	0.48437
Symbolism	0.552756
Symbolize	0.584136
Symbolic	0.550897
Symbols	0.645067
Simmer	0.532114
Cement	0.454848

The barcode duty cycle for the visual speech signals $VS_{BLI}^{uc}{}^{W_m}(f_i), VS_{BLI}^{lc}{}^{W_m}(f_i)$ and those, which are generated from visual speech signals obtained from the sample set $uclc^{W_m}[n]$, are compared in Figure 7-33.

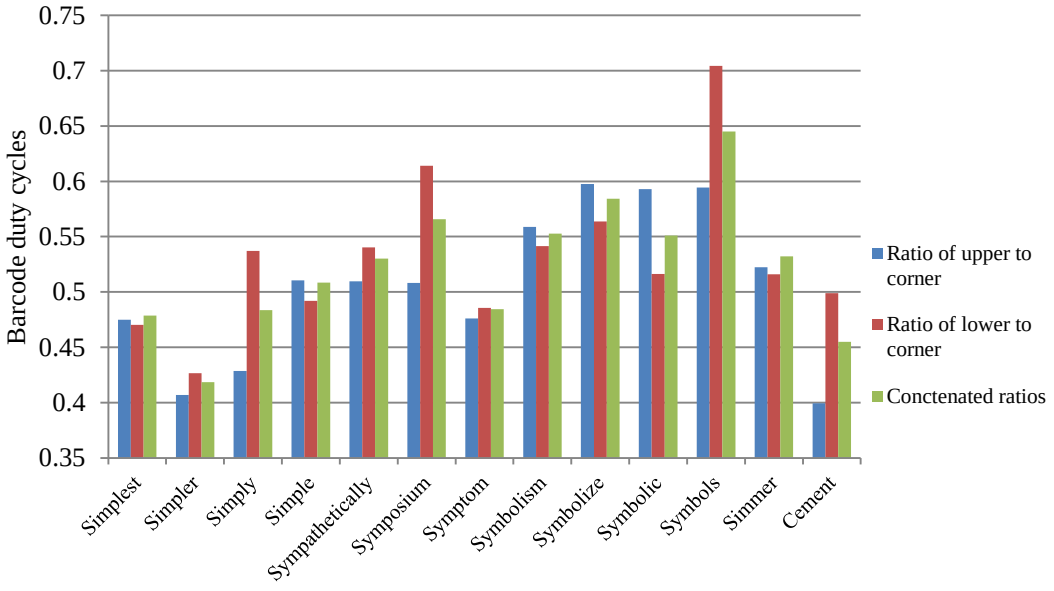


Figure 7-33: The barcode duty cycles for the visual speech signals

The visual words can be identified by the duty cycles of each barcode set however, they accuracy of identification should be high since they have similar duty cycles.

7.4. DIGITAL REPRESENTATION OF VISUAL SPEECH SIGNALS

The next approach uses a coding scheme to allocate a unique expression to the visual word. This unique expression is called digital visual word. Toward coding the visual speech sample sets, it is necessary to quantize the signal via partitioning the range of samples to different and limited number of levels. In this work, a coding scheme chosen as Huffman coding to derive a binary signature for the extracted feature sample sets. The quantized amplitudes are the coding symbols and represented by digits. In Figure 7-34, the schematic of components for allocating digital expressions to the visual word is depicted.

The provided visual speech signals from the normalized ratios of the upper to corner, lower to corner visual speech sample sets as well as their concatenation, are quantized. In parallel, the quantizing levels are determined via statistical characteristics of estimated visual speech signals density function. In the last stage, the Huffman coding scheme represents the quantized visual speech signals.

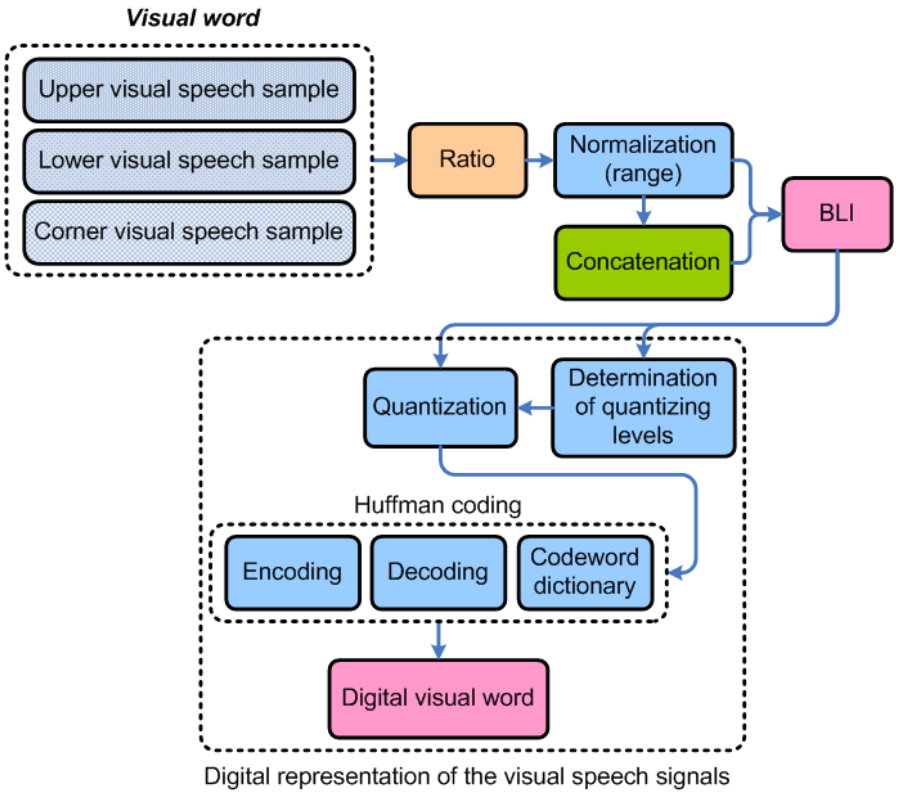


Figure 7-34: The block diagram of the digital word generation method

In the next two sections, first the method of defining quantizing levels are represented and in the second section the process of coding the quantized signals is described.

7.4.1. VISUAL SPEECH SIGNALS QUANTIZATION

In this section, before using the encoder, data quantization is used for compressing the data. The compression stage consists of reducing the amplitude levels to quantized and limited amplitude. Simplicity in data probability analysis is the main benefit of this operation on the original visual speech sample sets. Finding the quantizing level is the first stage for obtaining the digital representation for the visual speech signals. In Figure 7-35, the suggested process is represented where the distributions of the visual speech signals is the reference for determination of quantizing levels. The other two complimentary components are the statistical information obtained by the density estimation method as well as averaging density functions for defining a single reference for extracting the statistical information.

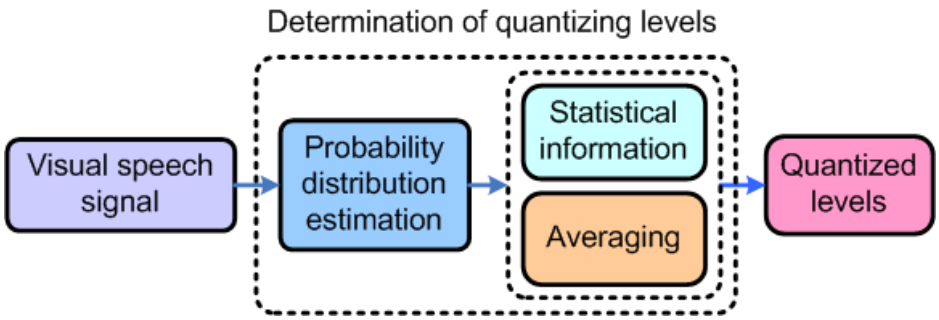


Figure 7-35: Determination of the quantizing levels from visual speech signals statistical information

For more efficiency in designing the coding scheme, the procedure, which is employed in statistical analysis in Section 5.4.2, is useful. This information about the distribution of amplitudes determines the maximum amplitude to be set as the maximum quantized level. According to the concentration of amplitudes in the distribution of original visual speech sample sets, the quantizing levels could be set for covering the maximum number of amplitudes. The output relation of a quantizer using the overall probability density function is depicted in Figure 7-36. The dashed lines are quantizing the input signal near to the peak of density function. In other words, the input signal samples would be appeared with higher probability near the peak of density function that needs to be quantized more accurately.

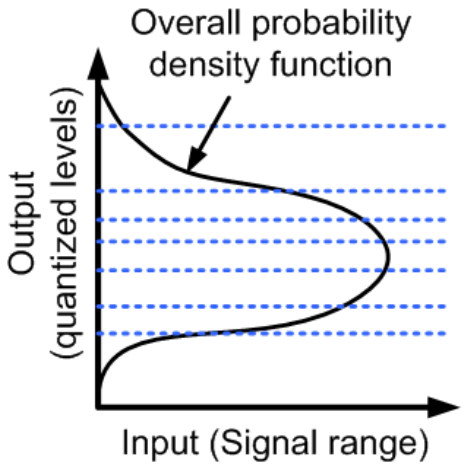


Figure 7-36: The effect of defining quantizing levels via density function

The visual speech signals amplitude after quantization process could be thought as a set of signals amplitudes with discrete values. In other words, the quantization method generates discrete-amplitude signals. More specifically, the variation of samples amplitudes can be dimensionally reduced using quantization method. The quantized visual speech signals are the fundamental stage to achieve the digital representation of visual words. In the MATLAB environment, the quantization process starts with defining quantizing partitions. As mentioned the partitions are dividing the range of input signal. In this work, partitioning performed using the density functions of visual speech signals $VS_{BLI}^{uc}{}^{W_m}(f_i)$, $VS_{BLI}^{lc}{}^{W_m}(f_i)$ and $VS_{BLI}^{uclc}{}^{W_m}(f_i)$ are averaged in order to construct a general density function that inherits a similar statistical property from all three types of visual speech signals. The averaged density function is denoted by A. The partitions are bounded between the maximum and minimum of the density function. The number of quantizing levels is selected $l = 8$. The levels are representing by a vector called codebook. The codebook assigns the quantizing values to the appropriate regions between petitions. Based on the codebook length, the partitions can be represented by a vector with length of 10. It means whenever a sample amplitude is located between a partition, it is mapped to a fixed eight possible values. To describe the procedure, recalling Eq. (2-34), the quantizing function is:

$$x_q[n] = \begin{cases} L_0 & if \quad P_1 = min(A) \le {VS_{BLI}}^{W_m}(f_i) \\ L_0 & if \quad P_1 = min(A) \le {VS_{BLI}}^{W_m}(f_i) < P_2 \\ L_1 & if \quad P_2 \le {VS_{BLI}}^{W_m}(f_i) < P_3 \\ . & . \quad . \\ . & . \quad . \\ . & . \quad . \\ L_7 & if \quad P_8 \le {VS_{BLI}}^{W_m}(f_i) < P_9 = max(A) \\ L_7 & if \quad P_9 \le {VS_{BLI}}^{W_m}(f_i) \end{cases} \tag{7-16}$$

for $f_i = [0, \beta(F_{W_m} - 1)]$. In Eq. (7-16) two partitions P_1 and P_9 has been determined as the minimum and maximum values of the averaged estimated density function A. The other partitions are also defined using the estimated density function. More specifically, the mean and standard deviation of the density function are used for defining the partitions. The frequency of samples in the estimated density function is often asymmetric. Therefore, an adjustment on the values of could be helpful by visual consideration of the density function. This will cause the more frequent samples to appear as peaks on the density function are included in a particular partition. This flexibility is denoted by $\epsilon_i,\ i = \{1, 2, 3, \dots, 9\}$ and is added to the partition values. In addition, the quantization levels are adjusted to obtain the signal between zero and one. In Table 7-5 the partitions and corresponding values are represented.

Table 7-5: Partition assignment using the estimated density function of the visual speech signals

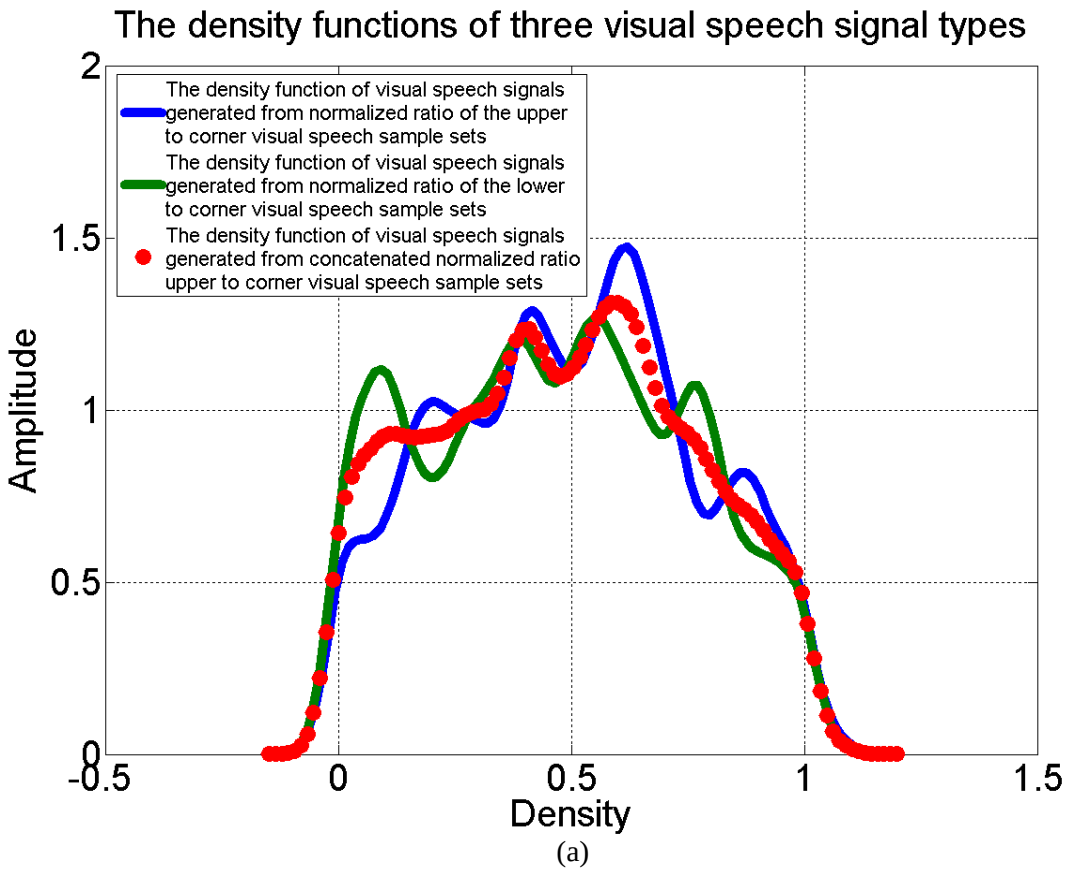
Quantizing levels	Partitions
$min(A) + \epsilon_1$	P_1
$\frac{mean(A) - std(A) + min(A)}{2} + \epsilon_2$	P_2
$mean(xI1) - std(xI1) + \epsilon_3$	P_3
$\frac{2mean(A) - std(A)}{2} + \epsilon_4$	P_4
$mean(A) + \epsilon_5$	P_5
$\frac{2mean(A) + std(A)}{2} + \epsilon_6$	P_6
$mean(xI1) - std(xI1) + \epsilon_7$	P_7
$\frac{mean(A) + std(A) + min(A)}{2} + \epsilon_8$	P_8
$max(A) + \epsilon_9$	P_9

After allocating values to the partitions, the quantizing levels are obtained by finding the geometrical mean of pairs of succeeding partitions. Therefore, the quantizing levels in Eq. (7-16) are:

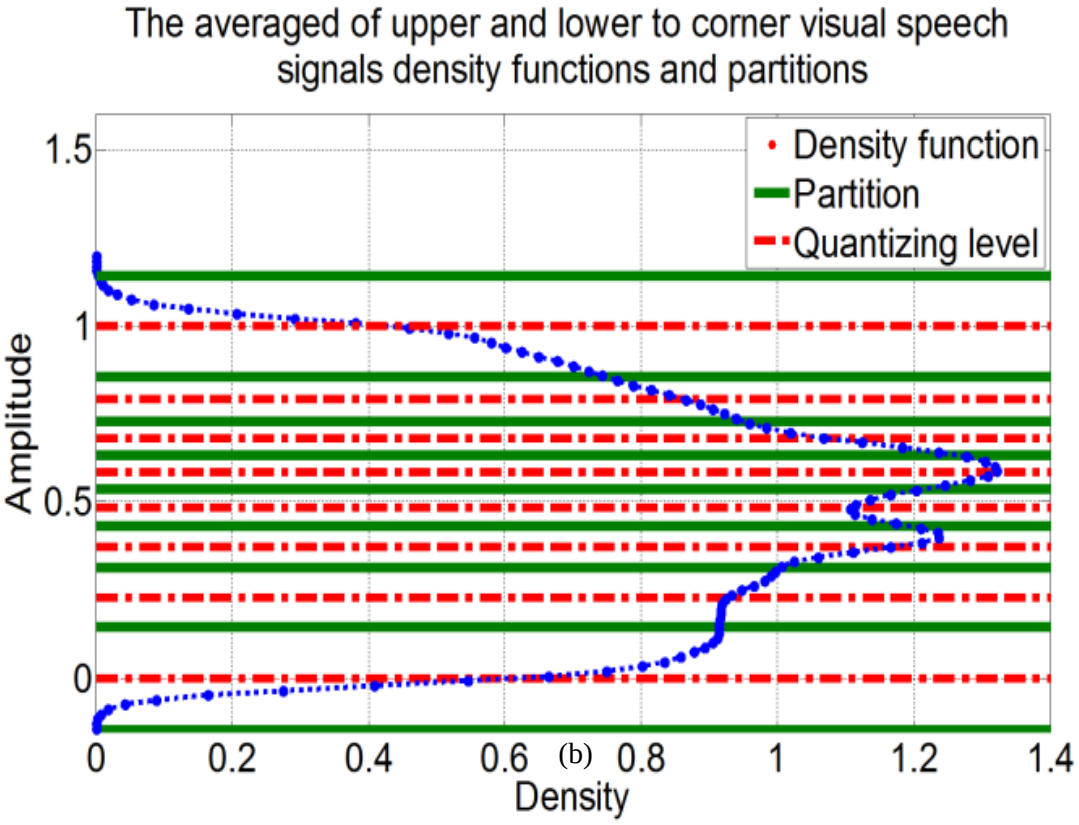
$$L_i = \begin{cases} L_0 = \dfrac{P_1 + P_2}{2} \\ L_0 = \dfrac{P_1 + P_3}{2} \\ L_1 = \dfrac{P_3 + P_3}{2} \\ \quad . \\ \quad . \\ \quad . \\ L_7 = \dfrac{P_8 + P_9}{2} \\ L_7 = \dfrac{P_8 + P_9}{2} \end{cases} \tag{7-17}$$

The quantization errors of the quantized visual speech signals and concatenated visual speech signals, are represented by $(VS_{BLI}^{uc\ W_m}/N^{uc})_q$, $(VS_{BLI}^{lc\ W_m}/N^{lc})_q$ and $(VS_{BLI}^{uclc\ W_m}/N^{uclc})_q$.

In this part, the visual speech signal (Section 7.1) quantization has been demonstrated. The results are consisted of quantized and the encoded visual speech signals. Using the density function estimation, the density of the visual speech signal $VS_{BLI}^{uc\ W_m}(f_i)$, $VS_{BLI}^{lc\ W_m}(f_i)$ and $VS_{BLI}^{uclc\ W_m}(f_i)$, $m = \{1,2,3,\dots,13\}$ has been determined, as shown in Figure 7-37 (a). According to Eq. (7-17) and manual adjustments represented in Table 7-6, the average density function of the three sample sets is partitioned, and illustrated in Figure 7-37 (b).



(a)



(b)

Figure 7-37: The density (a) estimation of visual speech sample sets and (b) the geometric positions of partitions on the averaged density functions

Table 7-6: The Adjustment values corresponding to Table 7-5

Values of ϵ							
$\epsilon_1 = 0.002$	$\epsilon_2 = 0.15$	$\epsilon_3 = 0.18$	$\epsilon_4 = 0.1$	$\epsilon_5 = -0.09$	$\epsilon_6 = -0.19$	ϵ_7=-0.2	ϵ_7=-0.053

In Table 7-7, the input-output values of the quantizer are represented and in Figure 7-38 these values are related to each other. The input signal is partitioned and the quantized signals appear in the output as combinations of the levels allocated by the codebook.

Table 7-7: The input-output relation of the quantizer

Partitions		Codebook	
No.	Amplitude	Level	Quantized amplitudes
1	-0.141		
2	0.146	1	0.001
3	0.314	2	0.230
4	0.430	3	0.372
5	0.536	4	0.483
6	0.632	5	0.584
7	0.728	6	0.680
8	0.857	7	0.793
9	1.142	8	0.999

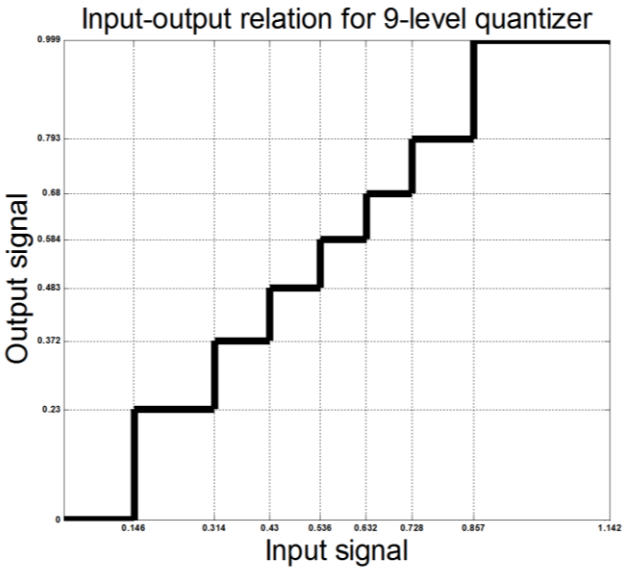


Figure 7-38: The input-output relation of eight levels quantizer

- **QUANTIZED VISUAL SPEECH SIGNALS** $\boldsymbol{VS_{BLI}^{uc}}^{W_m}[\boldsymbol{f_i}]$

In Figure 7-39, the visual speech signals $VS_{BLI}^{uc}{}^{W_m}(f_i), m = \{1,2,3,\dots,13\}$ the partitions and the quantized signals $VS_{BLI}^{uc}{}^{W_m}[f_i],\ m = \{1,2,3,\dots,13\}$ have been visualized.

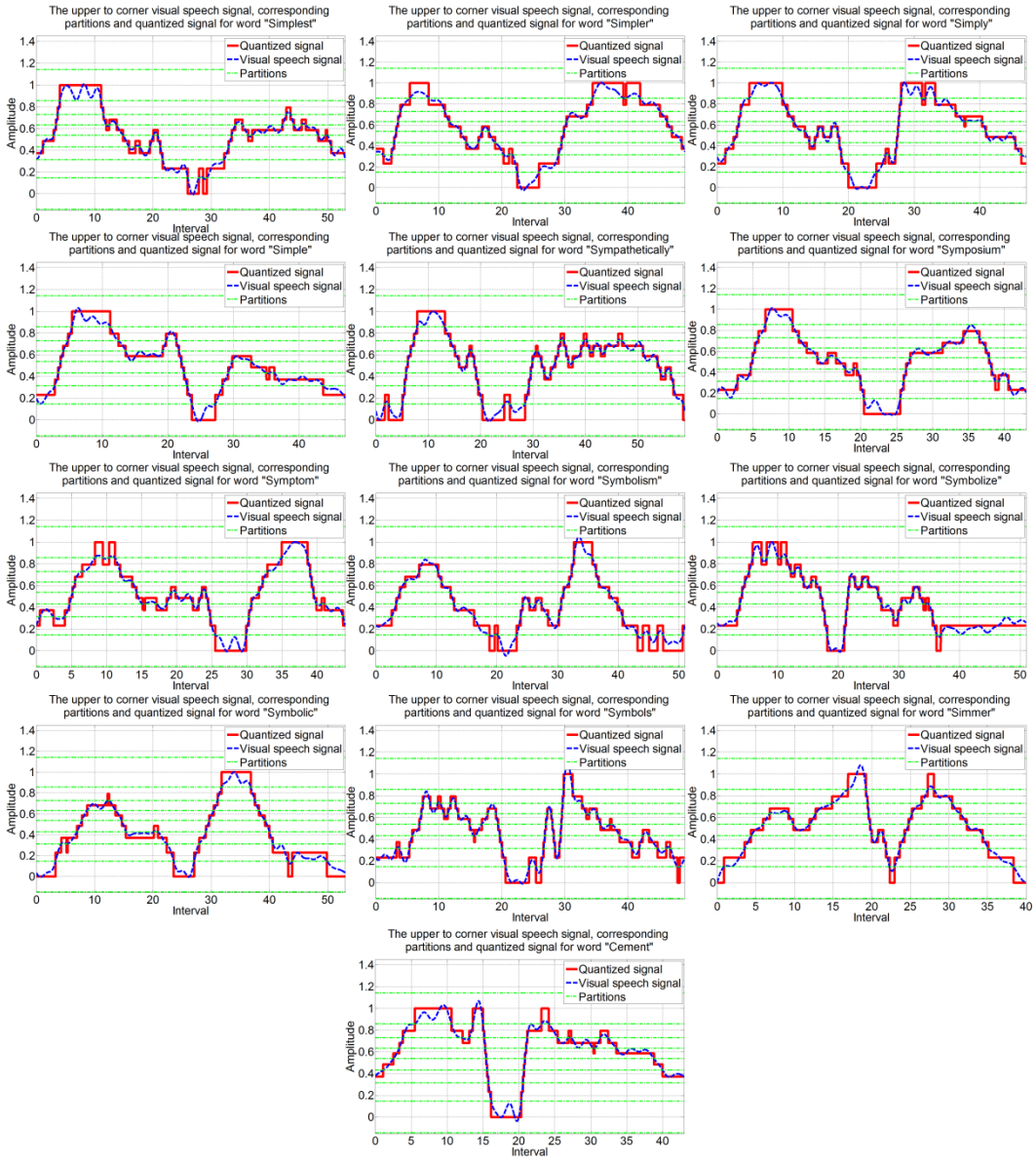


Figure 7-39: The quantized visual speech signals constructed by the normalized ratio of upper to corner visual speech sample sets for all selected words

- **QUANTIZED VISUAL SPEECH SIGNALS** $\boldsymbol{VS_{BLI}^{lc}}^{\boldsymbol{W_m}}[\boldsymbol{f_i}]$

In Figure 7-40, the visual speech signals $VS_{BLI}^{lc}{}^{W_m}(f_i), m = \{1,2,3,\dots,13\}$ the partitions and the quantized signals $VS_{BLI}^{lc}{}^{W_m}[f_i],\ m = \{1,2,3,\dots,13\}$ have been visualized.

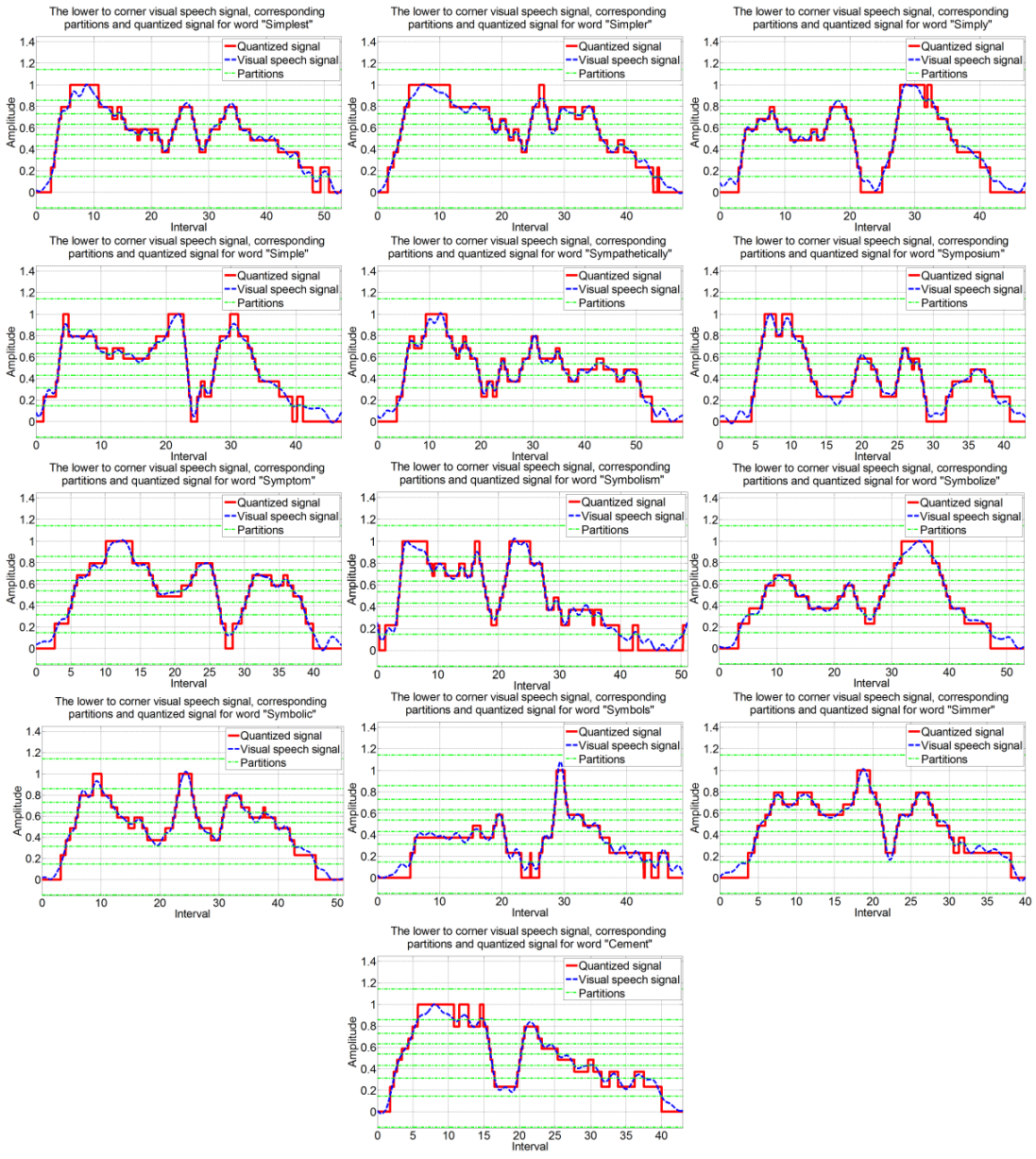


Figure 7-40: The quantized visual speech signals constructed by the normalized ratio of lower to corner visual speech sample sets for all selected words

Figure 7-39 and Figure 7-40 represent the quantized versions of visual speech signals shown Figure 7-10 and Figure 7-11. As mentioned in Section 2.14.1, the quantized

visual speech signals have discrete limited range of amplitudes. Therefore, the quantized visual speech signals are the simplified versions of the continues-time signals. It means the quantized signals consist of lower number of amplitudes which is selected by Eq. (7-17), whenever the transition occurs in the amplitude of continuous signal.

- **QUANTIZED VISUAL SPEECH SIGNALS** $\boldsymbol{VS_{BLI}^{uclc^{W_m}}[f_i]}$

In Figure 7-41, the visual speech signals $VS_{BLI}^{uclc^{W_m}}(f_i), m = \{1,2,3,\dots,13\}$ the partitions and the quantized signals $VS_{BLI}^{uclc^{W_m}}[f_i],\ m = \{1,2,3,\dots,13\}$ have been visualized.

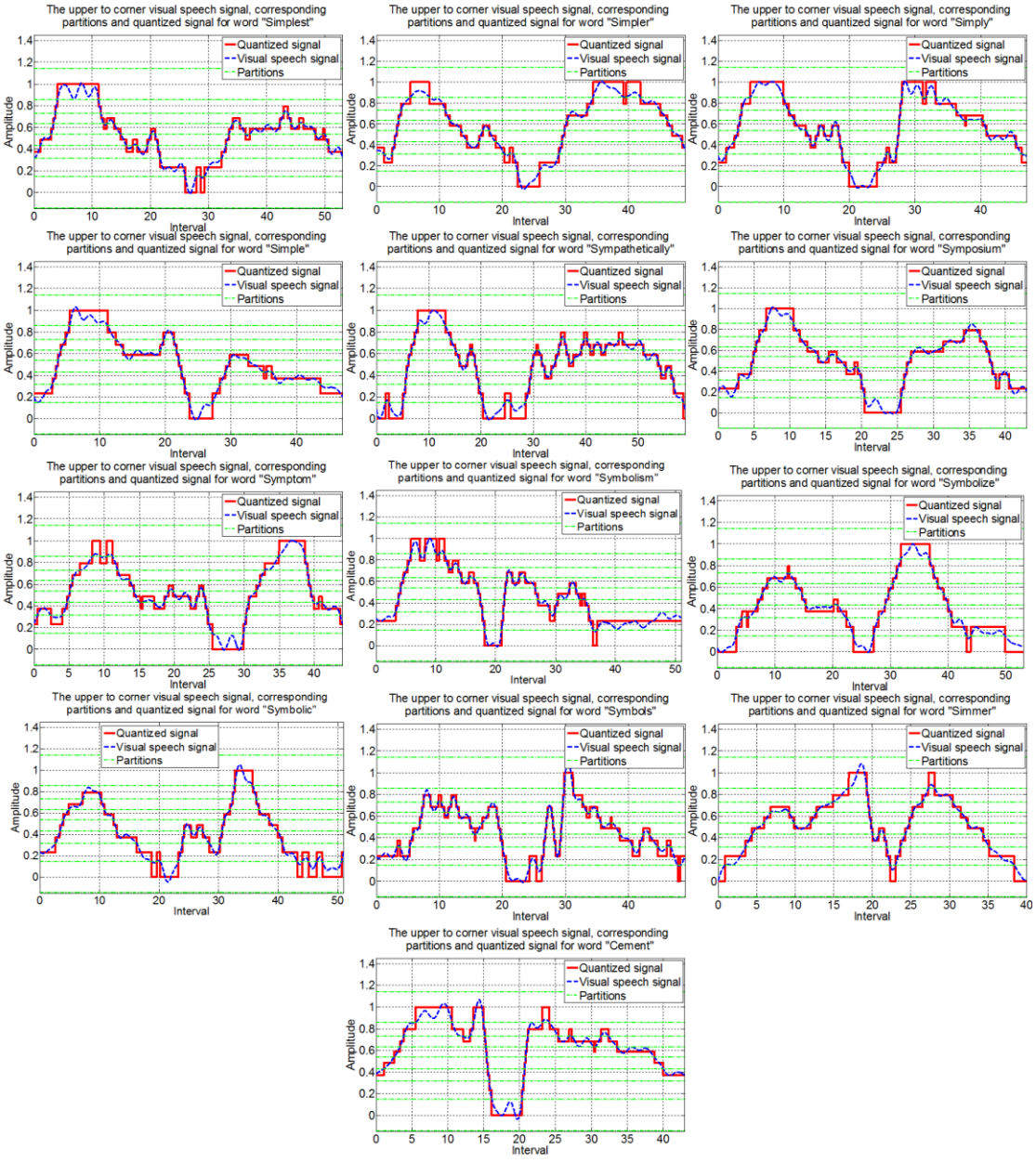


Figure 7-41: The quantized concatenated visual speech signals constructed by the normalized ratio of upper to corner and lower to corner visual speech sample, the carrier, and the PWM waveforms for all selected words

7.4.2. CODING THE VISUAL WORDS USING THE HUFFMAN ENCODING METHOD

After quantizing the visual speech signals, each level could be mapped to a limited set of values. These values could be a set of natural numbers or symbols $s_q, q = 0, 1, 2, \dots, l$. All samples are belonging to a group of values representing a particular symbol.

In the MATLAB environment, the Huffman coding has located in the communication toolbox. The corresponding function for defining a dictionary is called 'huffmandict'. The dictionary is generated using the symbols and their probabilities according to Table 7-8 to Table 7-15 Using such dictionary and the quantized signal, a function called 'huffmanenco' the coded signal would be generated. Recovering the quantized signal is also possible using the dictionary and the encoded signal using a function called 'huffmandeco'.

The quantized visual speech signals are represented by descript notation of their variable as $VS_{BLI}^{uc}{}^{W_m}[f_i]$, $VS_{BLI}^{lc}{}^{W_m}[f_i]$ and $VS_{BLI}^{uc}{}^{W_m}[f_i]$, $m = \{1,2,3,\dots,13\}$. In the coding scheme, the quantized visual speech signals are generated from a source (lip dimension). Since the priory knowledge about the nature of signals is provided, each quantizing level can be assumed as a symbol that occurs with a certain probability in complete set of signals. These symbols are used to allocate binary values called message codes to the quantized visual speech signals.

- **DIGITIZED VISUAL SPEECH SIGNALS** $DNRVS_{BLI}^{uc}{}^{W_m}$

In Table 7-9, the Huffman encoding information of the quantized visual speech signals $VS_{BLI}^{uc}{}^{W_m}[f_i]$, $m = \{1,2,3,\dots,13\}$ are represented. Each quantized level occurs with a probability p_i, $i = \{0, 1, 2, \dots, 7\}$. In Table 7-9, the massage codes and their lengths are represented. By multiplying the probabilities with the bit lengths, which are represented in Table 7-9, the frequency of each symbol (quantized level) is determined. The information content of each symbol is $I = -log_2(p_i)$. Consequently, the entropy of each symbol defines as the possibility of the symbol occurrence multiplied by its information. The less probable symbols convey more information and have higher entropy. This fact can be related to the less probable starting and

stopping amplitudes in the visual speech signals. The total entropy of codes is calculated by summation of symbol entropies.

Table 7-8: The Huffman coding characteristics of encoded the visual speech signals constructed from the normalized ratio of upper to corner visual speech sample sets

The Huffman encoding scheme properties for the quantized visual speech signals $VS_{BLI}^{uc}{}^{W_m}[x]$, $m = \{1,2,3,\dots,13\}$								
Message	0.0010	0.2308	0.3730	0.4839	0.5849	0.6809	0.7931	1.0000
Probability p_i	0.05950	0.13819	0.15930	0.17082	0.16506	0.14779	0.11132	0.04798
$p_i L_i$	0.23800	0.41458	0.47792	0.34165	0.49520	0.44337	0.33397	0.19193
$I = -log_2(p_i)$	4.07094	2.85521	2.6501	2.54940	2.59887	2.75835	3.16715	4.38128
$H = -p_i log_2(p_i)$	0.24222	0.39457	0.42218	0.43550	0.42898	0.40766	0.35258	0.21023
Total entropy	$H = 2.893961$							

Table 7-9: The Huffman code book for encoding the quantized visual speech signals $\boldsymbol{VS_{BLI}^{uc}{}^{W_m}[f_i]}$, $\boldsymbol{m = \{1,2,3,\dots,13\}}$

Symbol number i	Message Code	Bit length L_i
1	1010	4
2	011	3
3	001	3
4	11	2
5	000	3
6	010	3
7	100	3
8	1011	4

In Table 7-10, the digital visual speech signals obtained by the codebook in Table 7-9 for the quantized visual speech signals $VS_{BLI}^{uc}{}^{W_m}[f_i]$, $m = \{1,2,3,\dots,13\}$. The digital version is obtained by function called 'huffmanenco' in MATLAB environment.

Table 7-10: The digital visual speech signals generated from the quantized visual speech signals $\boldsymbol{VS_{BLI}^{uc}{}^{W_m}[f_i]}$, $\boldsymbol{m = \{1,2,3,\dots,13\}}$

Simplest	0110011100001010010111000100000100001100111001110001100101111010011101001100111000010000110000101000100000100001100011001011
Simpler	0010110011100001010010111000100001100111000110010111001011101001100111000010100101110010111000100000110011
Simply	0110011100001010010111000100001100111000110001100101111010011001011001110000101010010111001011100010000010000110010111
Simple	0111010011001110000101001011100010000010100010000110010111010011001110001100111001011
Sympathetically	1010011101001100111000010100101110001000011000010000110010111010011101001100111000010000110011100011000010100010000011000010100010000010000001010001000011001110100101111010
Symposium	0110011100001010010111000100001100011001110010111010011001110000101000100001100101100111011
Symptom	0110010110011100001010010111001011100010000110011100111000110011100011000110010

	11101001100111000010100101110001000011001110010 11
Symbolism	01100111000010100101110010111001011100010100010000010000110010111010011001 11000010000010000110010110011100011001110010111010011
Symbolize	10100110010110011100001010001000011010110010111010011001110000101001011100 01000011001011101001110 10
Symbolic	01100111000010100010000110010111010011101001100111001110010110011100001010 01011100010000110010111010011101001110 10
Symbols	01100101100111000010100010100010000010100010000110011100001000011001011101 00111010011001110000100001100101100111000010100101110001000001000011000110 01011001110010110010111010011
Simmer	10100110011100001000011000010100101110001000011001110010111010011001110000 10100101110001000011001011 1010
Cement	00111000010100101110001010010111000100001100101110100110011100001010010111 00101110001010001000001010001000011001

- **DIGITIZED VISUAL SPEECH SIGNALS** $DNRVS_{BLI}^{lc\ W_m}$

Table 7-11, the Huffman coding properties are represented for the quantized visual speech signals $VS_{BLI}^{lc\ W_m}[f_i], m = \{1,2,3,\dots,13\}$ where the Huffman dictionary is represented in Table 7-12.

Table 7-11: The Huffman coding characteristics of encoded the quantized visual speech signals $\boldsymbol{VS_{BLI}^{lc\ W_m}[f_x], m = \{1,2,3,\dots,13\}}$

The Huffman encoding scheme properties for the quantized visual speech signals $VS_{BLI}^{lc\ W_m}[f_i], m = \{1,2,3,\dots,13\}$								
Message	0.0010	0.2308	0.3730	0.4839	0.5849	0.6809	0.7931	1.0000
Probability p_i	0.07515	0.11899	0.14613	0.16492	0.16492	0.15657	0.12526	0.04801
$p_i L_i$	0.30062	0.35699	0.43841	0.49495	0.49478	0.46972	0.37578	0.19206
$I = -log_2(p_i)$	3.73395	3.07099	2.77459	2.60010	2.60010	2.67506	2.99699	4.3803
$H = -p_i log_2(p_i)$	0.28063	0.36544	0.40547	0.42882	0.42882	0.41885	0.37540	0.21032
Total entropy	$H = 2.913786$							

Table 7-12: The Huffman codebook for encoding the quantized visual speech signals $\boldsymbol{VS_{BLI}^{lc\ W_m}[f_i], m = \{1,2,3,\dots,13\}}$

Symbol number i	Message Code	Bit length L_i
1	1000	4
2	101	3
3	010	3
4	000	3
5	11	2
6	001	3
7	011	3
8	1001	4

The digital visual speech signals, which are generated from the quantized visual speech signals $VS_{BLI}^{lc\ W_m}[f_i], m = \{1,2,3,\dots,13\}$ are represented in Table 7-13.

Table 7-13: The digital visual speech signals generated from the quantized visual speech signals $\boldsymbol{VS_{BLI}^{lc}}^{\boldsymbol{W_m}}[\boldsymbol{f_i}], \boldsymbol{m} = \{\mathbf{1}, \mathbf{2}, \mathbf{3}, \dots, \mathbf{13}\}$

Simplest	1000101010000110010111001011001011001110001100011000010000110010110011100001000011001011001110000101011000101100011000
Simpler	1000101010000110010111001011001110011100011000010000110010111001011001110010110010111001110000100000101011000
Simply	1000101010000110010110011100011000110010110011100001010110000101010000110010111000101110010111000101100111000010101100010101010100001100010111001011100111000010101100011100001011000
Simple	1000101010000110010111001011001110011100101110010110011100001010110001010101010100001100101110010100110011100001010110001010101010100001100101110010100110011100001010110001010
Sympathetically	1000101010000110010110010111001011001110010110011100001010101010101000011000010000110010110011100111000010000110000101000001010110001000011001011001110011100011000010000110000101011000
Symposium	1000101010000110010111001011100101100111000010101010000110000101010100001100001010101000011001110001100001010110001010100000101011000
Symptom	1000101010000110010111001011001110001100101100111000010101100010101000011001110001100001010110001010100000101011000
Symbolism	1011000101010000110010111001011001011001011001011100101100111000010101010000110010111001011001110000100000101010101010101011000101100010100011001011100101100111000010000010101010101010101011000101100010100
Symbolize	1000101010000110011100001000011000010101010000110010111001011001110000101011000
Symbolic	1000101010000110010111001011001110001100001000011001011100101100111000010000110010111001011001110000100001100101110010110011100001000011001011100001100111000010101100010100
Symbols	1000101010000010000110000101011000101010000110010111001011001110000101011000010011000110000101011000
Simmer	1000101010000110010110010111001110010111001011001110000101010100001100101100111000101010100001100101100111000010101010101011000
Cement	1000101010000110010111001011100101110010110011100001010101000011001011001110000100000101010101010101011000

- **DIGITIZED VISUAL SPEECH SIGNALS** $\boldsymbol{DNRVS_{BLI}^{uclc}}^{\boldsymbol{W_m}}$

Finally, the quantized visual speech signals ${VS_{BLI}^{uclc}}^{W_m}[f_i]$ are the subject of encoding. The properties of the coding scheme are represented in Table 7-15 where the complimentary information about the Huffman dictionary is represented in Table 7-16.

Table 7-14: The Huffman coding characteristics of encoded the visual speech signals constructed from the concatenated normalized ratio of upper to corner lower to corner visual speech sample sets

The Huffman encoding scheme properties for the quantized visual speech signals $VS_{BLI}^{uclc^{W_m}}[f_i]$, $m = \{1,2,3,\ldots,13\}$								
Message	0.0010	0.2308	0.3730	0.4839	0.5849	0.6809	0.7931	1.0000
Probability p_i	0.06295	0.12796	0.15583	0.16821	0.16408	0.15170	0.12074	0.04850
$p_i L_i$	0.25180	0.38390	0.46749	0.33642	0.49226	0.45510	0.36222	0.19401
$I = -log_2(p_i)$	3.98961	2.96615	2.68194	2.57162	2.60747	2.72068	3.04998	4.36576
$H = -p_i log_2(p_i)$	0.25115	0.37957	0.41793	0.43258	0.42785	0.41273	0.36826	0.21175
Total entropy	$H = 2.901843$							

Table 7-15: The Huffman code book for encoding the quantized visual speech signals $\boldsymbol{VS_{BLI}^{uclc^{W_m}}[f_i]}$, $\boldsymbol{m = \{1,2,3,\ldots,13\}}$

Symbol number i	Message Code	Bit length L_i
1	1010	4
2	011	3
3	001	3
4	11	2
5	000	3
6	010	3
7	100	3
8	1011	4

Table 7-16: The digital the visual speech signals generated from the concatenated normalized ratio of upper to corner lower to corner visual speech sample sets

Simplest	00111000010100101110001000001000011001110011100011001011101001100111000010 00011000010000110001100111001011101001100111000010100101110001010001000011 00011100001110000101000100001100111000010100010000110010111010011101 0
Simpler	00101100111000010100101110001000011001110001100101100101110100110011100001 01001011100101110001000011001011101001100111000010100101110001000001000011 00011001110000101001011100010000010100010100010000110011100101110100111010
Simply	01100111000010100101110001000011001110001100011001011101001100101100111000 01010010111001011100010000110010111010011001110000101000100001100011000010 10001000011001011101001100111000010100101110010111000100001100101111010
Simple	01100111000010100101110001000001010001000011001011101001100111000110010111 01001100111000010100101110001000001000001000001010010111000100001100101110 10011001011001110000101001011100010000110010111010011101 0
Sympathetically	10100111010011001110000101001011100010000110000100001100101110100111010011 00111000010000110000101000100001100001010001000001000001010001000011001110 01011101001100111000010100010100101110001000001010001000011001011001011001 11000110011100000101000100000100001100111000110011100111010010111010
Symposium	01100111000010100101110001000011000110011100101110100110011100001010001000 01100101100101110100110011100001010010111001011100010000110010110011100011 00111000010000110001100101110100110011100101111000010100101110001000
Symptom	01100101100111000010100101110010111000100001100111000110011100000101110100 11001110000101001011100010000110011100101110100010011100001010010111000100 00110000101000100001100101110100110011100001 0
Symbolism	01100111000010100101110010111001011100010100010000100001100101110100110011 11000010000010000110010110011100011001110010111010011001011101001100111000 01010010111000101000101000101001011100010000110010110011100001010010111000 10000110011100101100101110100111010011

Symbolize	0110011100001010001000011001110010111010011001110000101001011100010000110010111101001110100110011100001000011001110001100101100111000010100101110001000011001011101
Symbolic	011001110000101000100001100101110100111010011001110011100101100111000010100101110001000011001011101001110100111010011101001100111000010100101110001000011000110011100001010010111000100001100111000010100010000010000110010110010111010
Symbols	011001011001110000101000101000100000101000100001100001000011001011101001110100110011100001000011001011001110000101001011100010000010000110001100101100111001011001011101001110100110011100111000110010111010011101001100111000010100101110001000011001011101001110
Simmer	101001100111000010000110000101001011100010000110011100101110100110011100001010010111000100001100101110100110011100001010001010001000001010010111000100000110010110011100001010001000011001011001011101
Cement	001110000101001011100010100101110001000011001011101001100111000010100101110001010001010001000011001110010111010011001110000101001011100101110010111000100000110010110011100001010001000011001110010110010111010

- **QUANTIZATION ERROR**

The quantization errors $({VS_{BLI}^{uc}}^{W_m}/N^{uc})_q$, $({VS_{BLI}^{lc}}^{W_m}/N^{uc})_q$ and $({VS_{BLI}^{uclc}}^{W_m}/N^{uc})_q$ are calculated for the quantized visual speech signals ${VS_{BLI}^{uc}}^{W_m}[f_i]$, ${VS_{BLI}^{lc}}^{W_m}[f_i]$ and ${VS_{BLI}^{uclc}}^{W_m}[f_i]$, $m = \{1,2,3,\dots,13\}$ by signal to error ratios as it represented in Eq. (2-36). In Figure 7-42, the quantization errors of visual speech signals ${VS_{BLI}^{uc}}^{W_m}[f_i]$,${VS_{BLI}^{lc}}^{W_m}[f_i]$, $m = \{1,2,3,\dots,13\}$ in comparison with ${VS_{BLI}^{uc}}^{W_m}(f_i)$, ${VS_{BLI}^{lc}}^{W_m}(f_i)$, $m = \{1,2,3,\dots,13\}$ are represented. The quantization error is minimized for the word 'symbols' while it is maximum for the word 'simpler'.

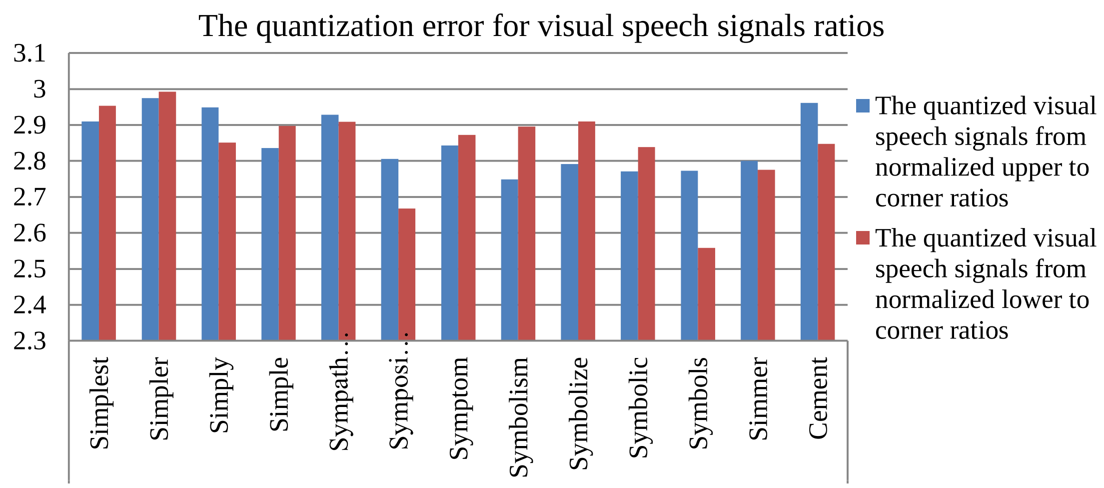


Figure 7-42: The quantization error between quantized and the non-quantized visual speech signals

Similarly, the signal to noise ratios for the quantized visual speech signals $VS_{BLI}^{uclc^{W_m}}[f_i], m = \{1,2,3,\dots,13\}$ are shown in Figure 7-43 The minimum and maximum errors belong to the words 'symbols' and 'simpler', respectively.

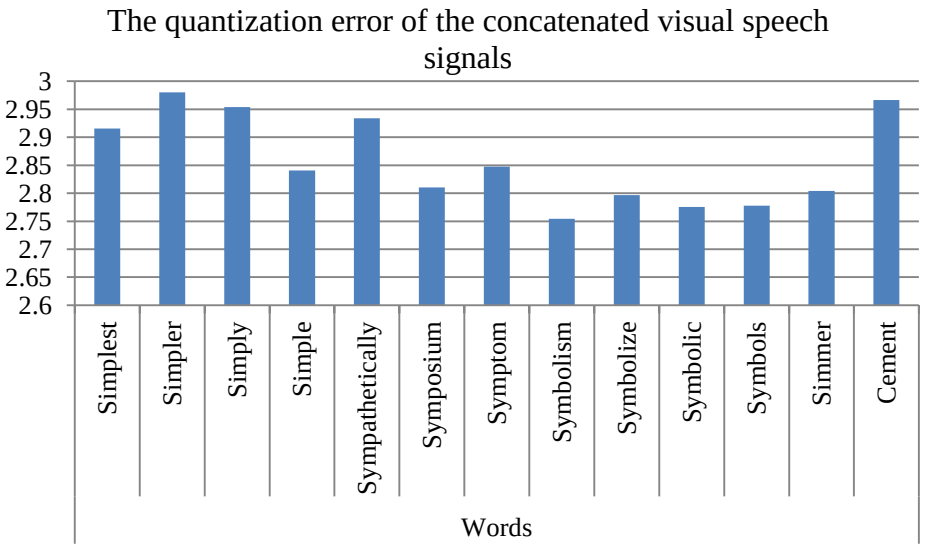


Figure 7-43: The quantization error between quantized and the non-quantized concatenated visual speech signals

Comparing the range of quantization errors in Figure 7-42 and Figure 7-43 reveals the higher errors occurring in the quantized visual speech signals $VS_{BLI}^{uclc^{W_m}}[f_i], m = \{1,2,3,\dots,13\}$.

7.5. THE 2D AND 3D VISUAL WORD SIGNATURES

Assume an invertible function $y = f(x), y \in \mathbb{R}, x \in \mathbb{N}$ describes the relation between an equidistance samples x and their amplitude y. Now, let another sample set function could be defined on the same domain as the function y is called $k = g(x)$. If inverse of it $x = g^{-1}(k)$ is substituted in the first function, the relation of the amplitudes in y and k will be obtained in composite form as$y = f\left(g^{-1}(k)\right)$. In the composite function the common interval x has been excluded. In other words, the domain and range of this function belongs to the samples in k and y, respectively. Most of the time and especially applicable to this work, it is very difficult to calculate the inverse of function (if they are invertible) or formula substitution to get$y = f\left(g^{-1}(k)\right)$. It is suitable to use a simple interfacing method for observation of the composite function.

By mapping the sample set in one of the function to the x-axis and the other to the y-axis in Cartesian system, such relation can be modelled. The obtained graph shows the proper result as if the composite functions of two sample sets or polynomials are calculated. Therefore, the 2D signatures of visual words with two graphs are:

$$F_{2D}{}^{uc_i} = (VSS_{BLI}^{u}{}^{W_m}(f_i), VSS_{BLI}^{c}{}^{W_m}(f_i)) \tag{7-18}$$

$$F_{2D}{}^{lc_i} = (VSS_{BLI}^{l}{}^{W_m}(f_i), VSS_{BLI}^{c}{}^{W_m}(f_i)) \tag{7-19}$$

The starting and stopping coordinates (end points) on the signature plot are coinciding. Therefore, a closed curve would be determined.

$$F_{2D}{}^{ul_0} = F_{2D}{}^{ul_{N-1}} \tag{7-20}$$

The C^0 continuity (Section 2.9.2) does not preserve in closed curves since the tangents in the endpoints are not collinear. Allocating a single 2D expression to visual word is possible. Using the expressions $BL(N(R(US, CS)))$ and $BL(N(R(LS, CS)))$, a scattering plot, which uniquely represents a signature for visual words, is obtained. The other category of allocating single representation to the visual word can be demonstrated by 3D visualization of upper, lower and corner visual speech signals. In this section, the concept of allocating a single expression to visual words brought up to the level of visualization. The relation graphical relation of upper, lower and corner visual speech signals are storing in a subgroup of visual word database called visual word 2D and 3D signature. The 2D signature has two branches. In one branch a visual word has two 2D graphical representations are obtained by allocating x axis to upper and lower visual speech signals and y axis to the corner visual speech signals. Therefore, there are two relations in a graph. In the next branch, the visual word has a single 2D representation derived from the BLI of normalized ratio of upper to corner visual speech sample sets and the BLI of normalized ratio of lower to corner visual speech sample sets. The other member in this family of database is represented by the 3D relation of upper, lower and corner visual speech signals for a word. This type of representation, which is more related to the perception of human, associates the concept of signatures. In the end of this section, the visual speech signals have unique signatures.

- **THE 2D SIGNATURES OF THE VISUAL SPEECH SIGNALS**

The visual speech signals $VS_{BLI}^{u}{}^{W_m}(f_i)$, $VS_{BLI}^{l}{}^{W_m}(f_i)$ are shown on y axes versus $VS_{BLI}^{c}{}^{W_m}(f_i)$, $m = \{1,2,3,\dots,13\}$ that are allocated to the x axes in Figure 7-44.

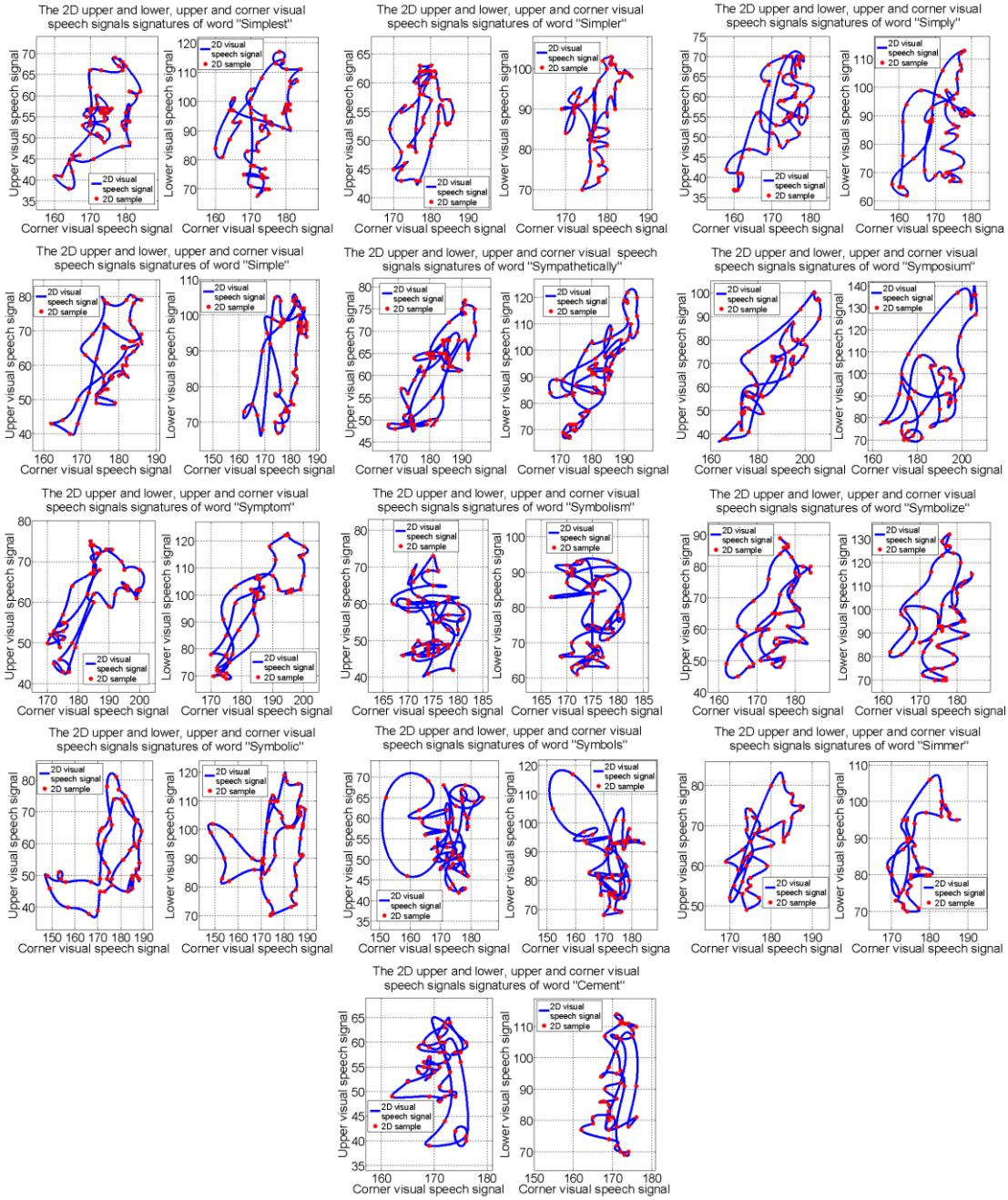


Figure 7-44: The 2D visual speech signatures of visual speech signals constructed by upper, lower and corner visual speech signals

- **THE 3D PLOTS OF THE VISUAL SPEECH SIGNALS**

In Figure 7-45, the visual speech signals $VS_{BLI}^{u}{}^{W_m}(f_i)$, $VS_{BLI}^{u}{}^{W_m}(f_i)$ and $VS_{BLI}^{u}{}^{W_m}(f_i)$, $m = \{1,2,3,\dots,13\}$ are visualized in 3D space by allocating y, z and x axis to them, respectively.

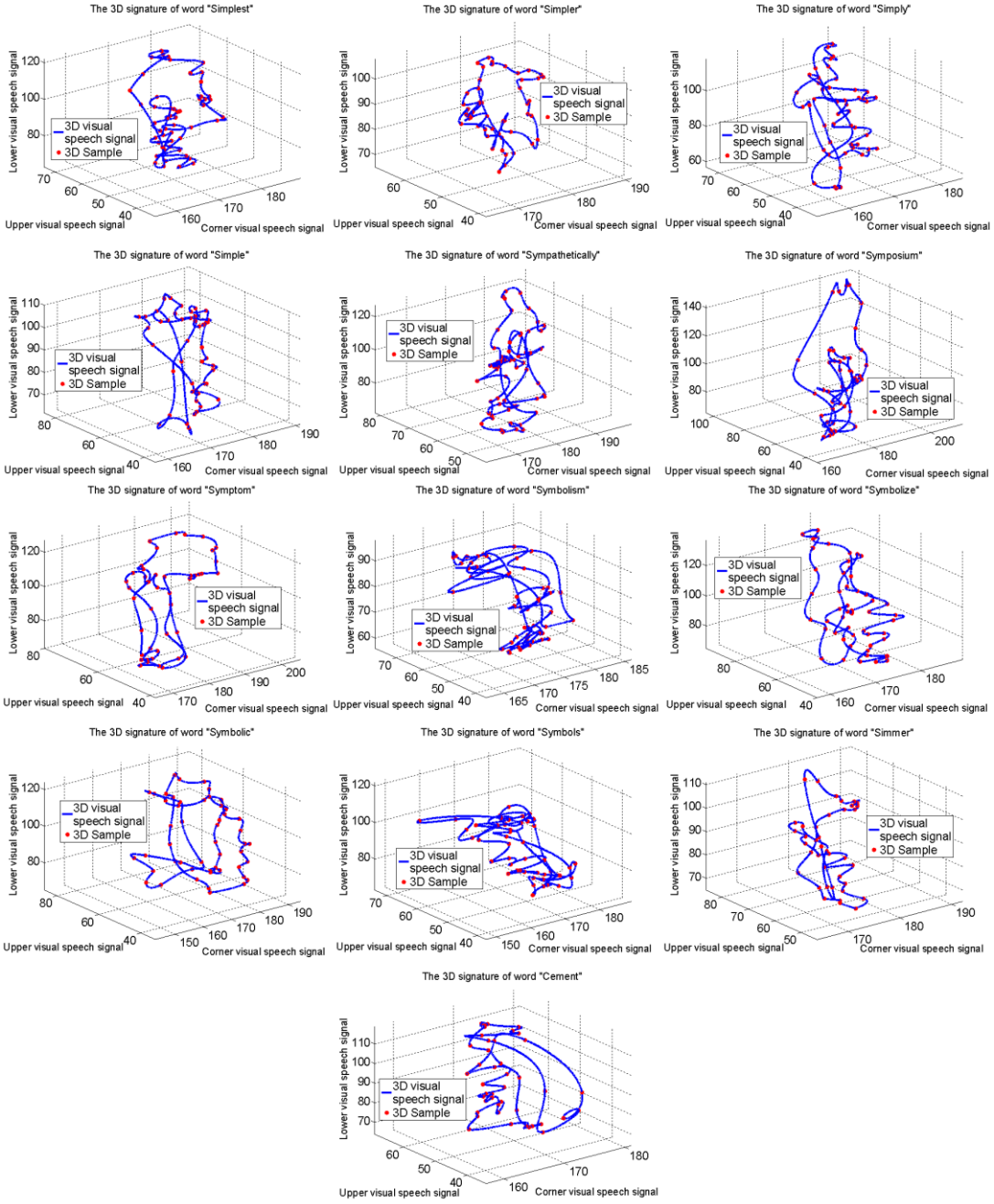


Figure 7-45: The 3D visual speech signatures of visual speech signals constructed by upper, lower and corner visual speech sample sets

- **VISUAL SPEECH SIGNALS 2D SIGNATURES**

In this section, the visual word is characterized by a single signature representation. Defining such signature needs x and y axes to be allocated to the visual speech signals $VS_{BLI}^{uc}{}^{W_m}(f_i)$ and $VS_{BLI}^{uc}{}^{W_m}(f_i)$, $m = \{1,2,3,\dots,13\}$ as it shown in Figure 7-46.

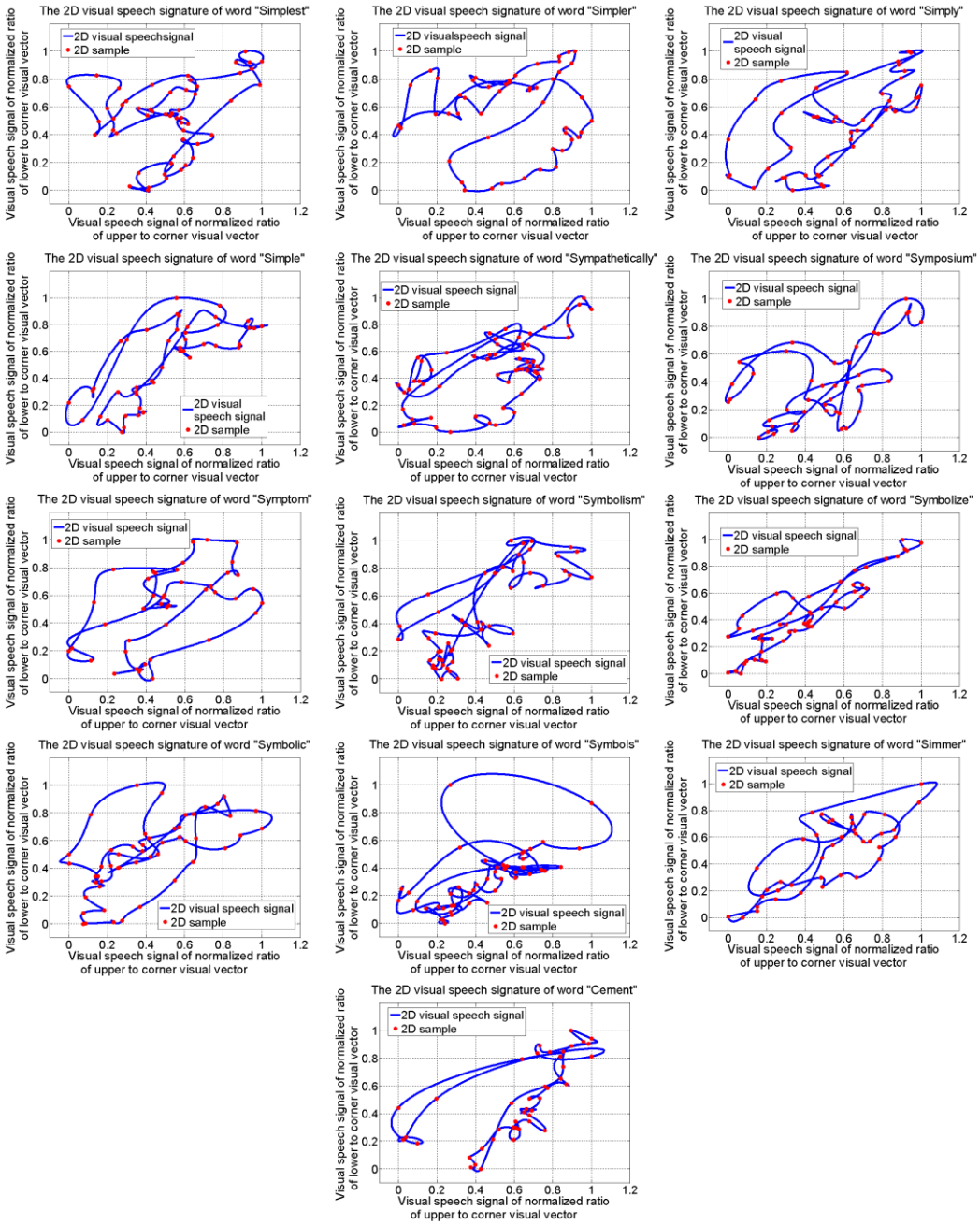


Figure 7-46: The 2D visual speech signatures of normalized visual speech signals constructed by the normalized ratio of upper to corner lower to corner visual speech sample sets for all selected words

The visual speech signatures in Figure 7-45 are visualised by allocating the x axis to corner visual speech signal and allocating the y axis to upper and lower visual speech signals, seperatly. Visualising the 3D signatures in Figure 7-46 are achieved by allocating the x axis to corner, y axis to upper and z axis to lower visual speech sample sets. The signuters are representing close trajectory curves in 3D domain. The final 2D visual speech signatures in Figure 7-47 are generated by allocating the normalized ratio of upper to corner visual speech signals in Figure 7-10 to x axis and the normalizes ratio of lower to corner visual speech signals in Figure 7-11 to y axis. This version of signature unifies the properties of lip dimensions during articulation of each word.

7.6. VOLUMETRIC MODELLING OF THE VISUAL WORDS

Previously, the visual speech sample sets processed for generating visual word signatures and its subgroups. In this section, the geometry of lip is modelled by the visual speech sample sets. On each frame, modelling the configuration of feature points is possible. The feature points in the coordinate of the ROI are located on the upper, lower and corner sides of outer lip's contour. In the most cases the lip geometry is symmetric vertically but it is not horizontally. The ROI can be translated to into Cartesian coordinates system. Therefore, the modified upper and lower visual speech sample sets are collinearly located on the y axis and consequently, the modified corner visual speech sample sets lies on the x axis. The corner sides of the lip, regarding to the centre point of mouth move to sideways with equal amounts while the upper and lower lips are moving upward and downward with different values during utterance. Formulating the geometrical relation between the feature points would associate a deformable lip template as well as providing the fundamentals for volumetric representation of visual words.

In the first step, the Bezier curve was adopted to define the simplest form of closed curve that could include the pairs of feature points. After finding such closed curves, very close similarities between parametric curves and sinusoidal curves is observed. By eliminating the parameter an ellipsoid form of Bezier closed curve is appeared. Therefore, the idea of including feature points to each other via ellipses is emerged. A mathematical expression for mouth outer contour with curve fitting and polynomial

methods as a form of template models can be obtained. Since mouth outer contour can be approximated with a closed curve, the Bezier method is applicable. The Bezier curve for rectangular arrangement of control points covering the ROI generates an ellipsoid form without symmetry related to rectangle. To tackle this imperfection, it is possible to define different curves by different symmetries and then average them. The ellipses are analysed into sinusoidal functions (Dwight, 1957) and are averaged. The other way is to adjust the control points in a way that the resulted curve shows an adjusted and symmetric ellipse to the ROI rectangle. In Figure 7-47 (a) the asymmetric resulted curve is shown while by adjusting the control points a symmetric closed curve with deferent covering degree is shown in Figure 7-47 (b).

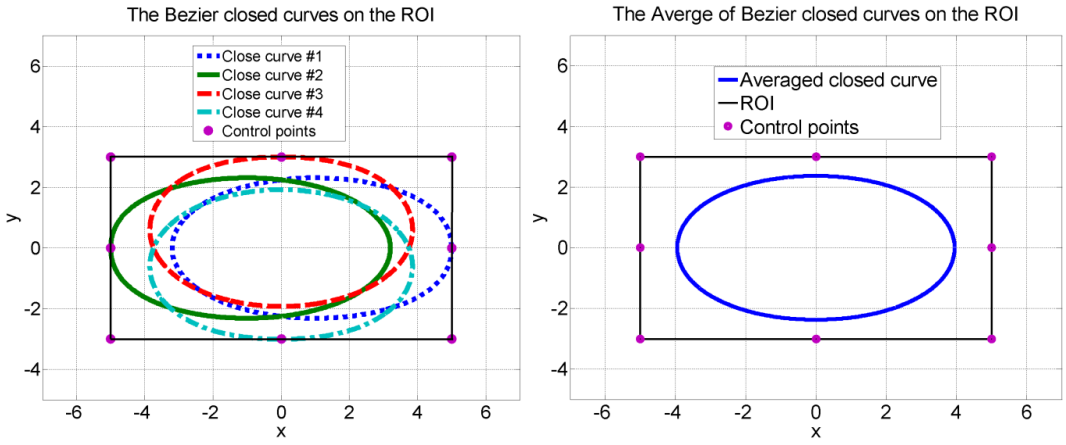


Figure 7-47: Four (a) Bezier closed curves and (b) their average closed curve on the ROI

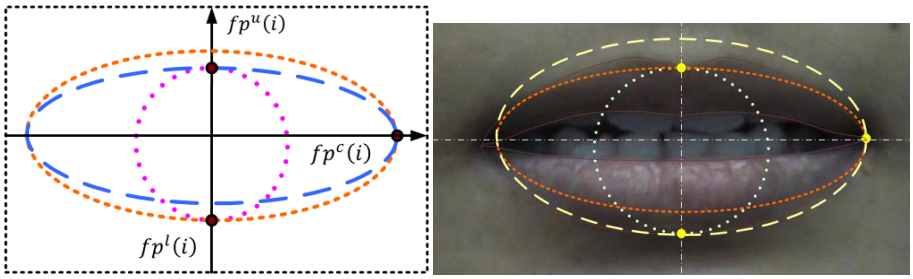
A convenience way to unify the Bezier closed curve that satisfies the symmetry of mouth is to find the approximation by ellipse equation. It is also possible to relate parameters of the Bezier curve to each other to get one expression in Cartesian coordinate but computational complexity causes serious time consumption to find complex solution that would practically inefficient.

In this section, the general formulation of mouth movement as a generic mathematical template is formulated by combination of ellipse formula and the BLI approach. This approach facilitates the dynamics of lip as it progress by frames in 3D space. The lip dynamics are trace by feature points where expressed by the BL polynomials as visual speech signals. In the same time, the feature points in each frame are also related by ellipses. The ellipses are located in the $z-y$ plan. There dimensions are changing

according to the provided visual speech sample sets on x axis making a 3D visualization of words.

- **LIP'S ELLIPSOID MODEL**

There are three possibilities of relating the feature points on each frame with each other. In Figure 7-48, the feature points are related via three ellipses.



(a) (b)

Figure 7-48: The model of lip (a) consisting of three ellipses is including visual features and (b) their location on an image ROI

The ellipse, which includes the upper and corner feature points, is similar to one that includes the lower and corner feature points. However, the asymmetric lip geometry does not provide equal ellipses. The upper and lower ellipsoid relation also defines circle since these feature points are co-linear. An ellipse can be expressed mathematically by:

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}=1 \tag{7-21}$$

For example, the upper $FP^u_{W_m}(j)$ and corner $FP^c_{W_m}(j)$ visual speech sample sets can be substituted in Eq. (7-21). For simplicity of expressions, the collection of ellipses is substituted with DT representations of the visual speech sample sets:

$$\frac{x^2}{u^{W_m}[n]^2}+\frac{y^2}{l^{W_m}[n]^2}=1 \tag{7-22}$$

$$\frac{x^2}{u^{W_m}[n]^2}+\frac{y^2}{c^{W_m}[n]^2}=1 \tag{7-23}$$

$$\frac{x^2}{l^{W_m}[n]^2}+\frac{y^2}{c^{W_m}[n]^2}=1 \tag{7-24}$$

or equivalently:

$$u^{W_m}[n]^{-2}x^2 + l^{W_m}[n]^{-2}y^2 = 1 \quad (7\text{-}25)$$

$$u^{W_m}[n]^{-2}x^2 + c^{W_m}[n]^{-2}y^2 = 1 \quad (7\text{-}26)$$

$$l^{W_m}[n]^{-2}x^2 + c^{W_m}[n]^{-2}y^2 = 1 \quad (7\text{-}27)$$

By replacing the visual speech signals into the sequence of ellipses formula, it is possible to also include the Lagrange polynomials and ellipse representation as:

$$x^2 {VSS^u_{BLI}}^{W_m}(f_i)^{-2} + y^2 {VSS^c_{BLI}}^{W_m}(f_i)^{-2} = 1 \quad (7\text{-}28)$$

$$x^2 {VSS^u_{BLI}}^{W_m}(f_i)^{-2} + y^2 {VSS^l_{BLI}}^{W_m}(f_i)^{-2} = 1 \quad (7\text{-}29)$$

$$x^2 {VSS^l_{BLI}}^{W_m}(f_i)^{-2} + y^2 {VSS^c_{BLI}}^{W_m}(f_i)^{-2} = 1 \quad (7\text{-}30)$$

where $x \in [0: F_{W_m} - 1]$. Eq. (7-28) to Eq. define another mathematical representation of lips feature points deformation that has been related to each other by three ellipses. As the lip pronounces a word, the ${VSS^u_{BLI}}^{W_m}(f_i)$, ${VSS^l_{BLI}}^{W_m}(f_i)$ and ${VSS^c_{BLI}}^{W_m}(f_i)$ polynomials are progressing and a volumetric representation of lip's dynamism is obtained. It is important to mention that in this representation of the lip's movements, the movement of lip's upper and lower contours are considered symmetric.

The volumetric representations of representing the visual words belongs to modelling the upper $u^{W_m}[j]$, lower $l^{W_m}[j]$ and corner $u^{W_m}[j]$, $j = \{0, 1, 2, \dots, F_{W_m} - 1\}$ visual speech sample sets by ellipses. In order to unify the lip counter template, the three ellipses in Figure 7-48 are averaged. Referring to Eq. (2-4), the averaging is conducted on trigonometric functions. The sequence of these ellipses in each visual word is rendered in the MATLAB environment, using the function 'surf'. These 3D shapes are called the volumetric representation of the visual words.

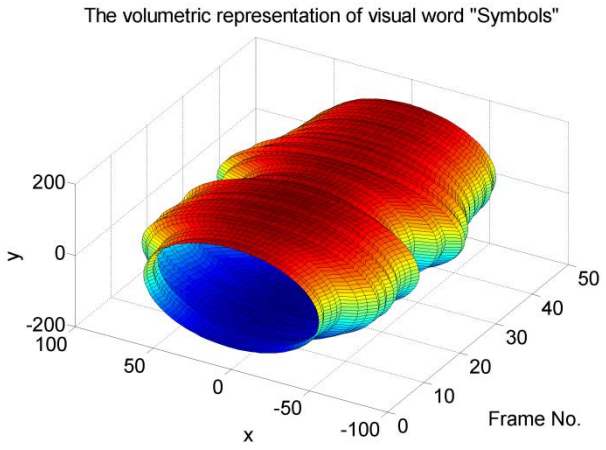


Figure 7-49: The volumetric representation of the visual word ‘symbols’

The ellipsoid lip template is used for volumetric representations of the visual words as it shown in Figure **7-49** for the visual word ‘symbols’. The average of these ellipses then is used for volumetric representation of a visual word. In Figure 7-50, the results are illustrated.

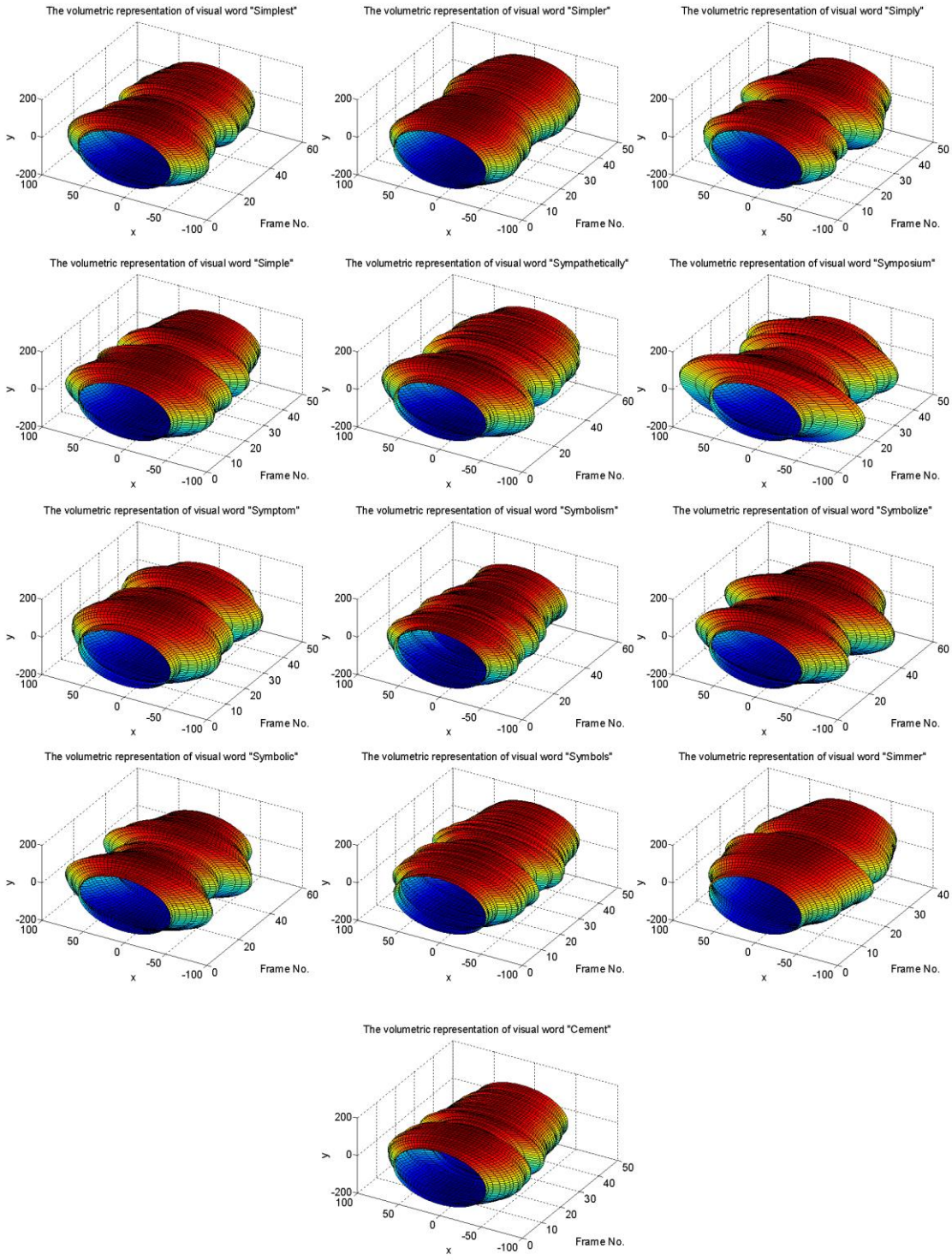


Figure 7-50: The volumetric representation of the visual words

In this part, the details of signature driven analysis of visual speech signals have been proposed in Figure 7-51. The same notations used in Figure 7-7, is also used:

- The concatenated visual speech signals are denoted by two brackets [.].

- The upper, lower, and corner lip's visual samples set are noted by US, LS and CS, respectively.
- Normalization is noted by N.
- The BL indicates the Barycentric Lagrange polynomials.
- The ratio of sample sets (US, LS and CS) are noted by R.
- The barcode generation is denoted by operation $B(.)$.
- The quantization and Huffman coding operations are denoted by $HC(.)$.
- The scattering representation of the visual speech signals are denoted by $S(.)$.

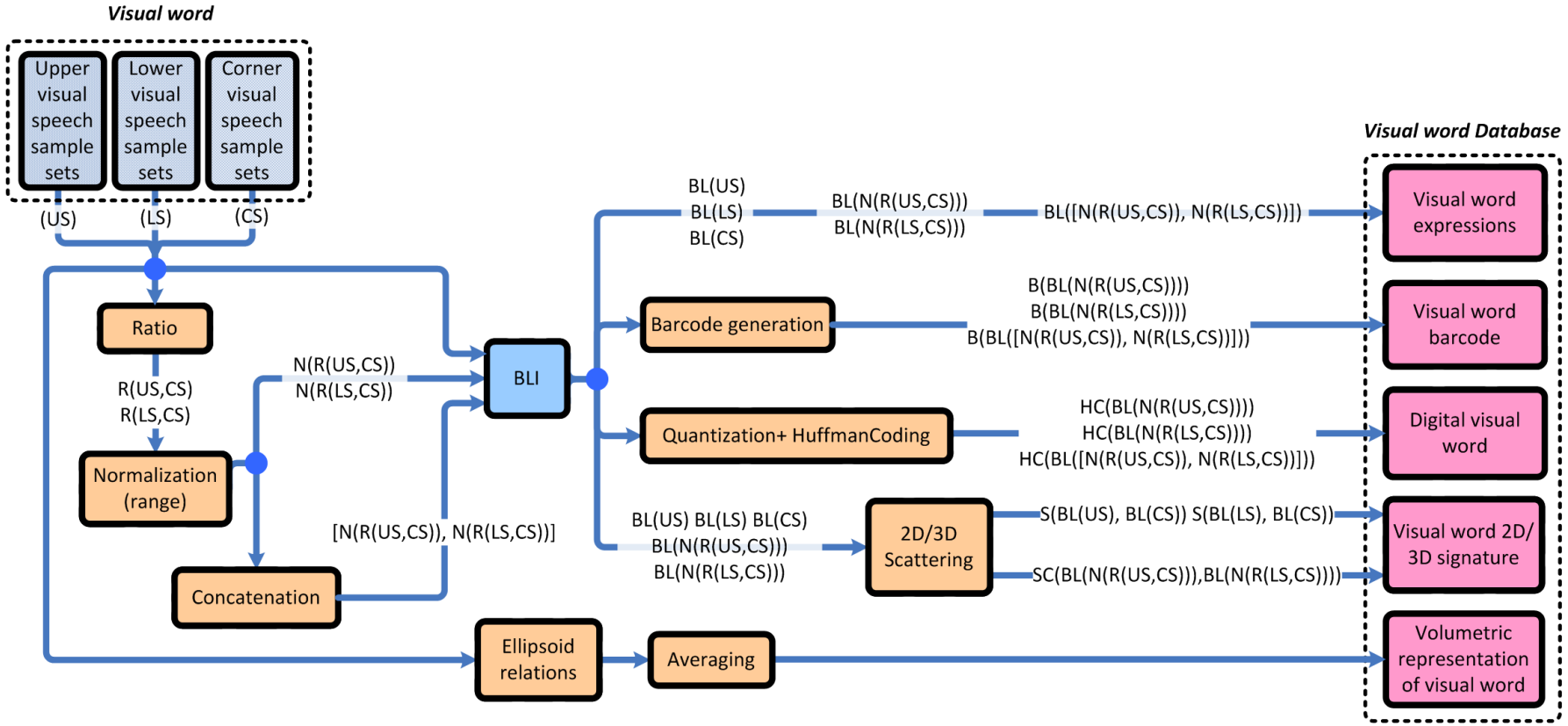


Figure 7-51: The overall schematic model of visual signature-based database method

The results of the operations are stored in visual word database. The results are categorized into three main families. The first family is provided by the mathematical expressions of the visual speech signals. These mathematical expressions include the visual speech signals driven by the upper, lower, and corner visual speech sample sets, the visual speech signals driven by the normalized ratio of the upper to corner and lower to corner visual speech sample sets. The second family of the information in the visual word database provides visualized representations of the visual speech signals and their geometrical properties. The first family members of this group are the visual speech signals which are driven by the upper, lower and corner visual speech sample sets,

The second family members are the visual speech signals driven by the normalized ratio of the upper to corner and lower to corner visual speech sample sets. The third family members in this group are the visual speech signals driven by the concatenation of the normalized ratio of the upper to corner and lower to corner visual speech sample sets. The forth family members in this group consists of the 2D representations of the visual speech signals driven by the upper versus corner and lower versus corner visual speech sample sets. The fifth family members in this group consists of the 3D representation of the visual speech signals driven by the upper, lower and corner visual speech sample sets. The unified visual word signatures defined by the visual speech signals driven by the normalized ratio of the upper to corner versus the visual speech signals driven by the normalized ratio of the lower to corner visual speech sample sets. Similarly, the barcode of the visual words is defined for the visual speech signals driven by the normalized ratio of the upper to corner and lower to corner visual speech sample sets as well as the visual speech signals driven by the concatenation of the normalized ratio of the upper to corner and lower to corner visual speech sample sets. In the third family, the visual speech signals are represented digitally. The digital representations are dedicated to the visual speech signals driven by the normalized ratio of the upper to corner and lower to corner visual speech sample sets as well as the visual speech signals driven by the concatenation of the normalized ratio of the upper to corner and lower to corner visual speech sample sets.

7.7. CONSTRUCTION OF SIGNAL USING THE SINC FUNCTION

It is possible to use the visual speech sample sets for signal construction by adopting the sinc function as the basic function of interpolating polynomial. The sample set and the discrete-time sample set $x[k] = \{a_0, a_1, a_2, \dots, a_{N-1}\}$ are directly related. In other words, the signal samples are defined in continues-time domain with impulse train where can be represented by discrete-time samples $x[f_i]$. Therefore, the constructed signal inherits the continuity from sinc function that is multiplied by each sample.

Therefore, Eq. (4-14) to (4-16) can be interpreted as multiplication of each sample's amplitude to the sinc function where is centred in the sample and summation of all multiplications. By allocating a high sampling rate to the sinc function, the resulted signals obtained in continuous form. The interpolating values between each span are defined by the summation of samples on the sinc functions. This type of signal construction also provides a mathematical representation of data sample that is suitable for formulating the speech visual speech sample sets. By substituting the discrete-time representation of visual speech sample sets from Eq. (4-14) to (4-16) into Eq. (2-26) the constructed signals are defined as:

$$VS_{sinc}^{u}{}^{W_m}(f_i) = \sum_{i=0}^{F_{W_m}-1} u^{W_m}[i]\, sinc(f_i - iT_0) \tag{7-31}$$

$$VS_{sinc}^{l}{}^{W_m}(f_i) = \sum_{i=0}^{F_{W_m}-1} l^{W_m}[n]\, sinc(f_i - iT_0) \tag{7-32}$$

$$VS_{sinc}^{c}{}^{W_m}(f_i) = \sum_{i=0}^{F_{W_m}-1} c^{W_m}[i]\, sinc(f_i - iT_0) \tag{7-33}$$

The visual speech signals driven by the normalized ratio of upper to corner and lower to corner visual speech sample sets are:

$$VS_{sinc}^{uc}{}^{W_m}(f_i) = \sum_{\square=0}^{F_{W_m}-1} N_{uc}{}^{W_m}[j]\, sinc(f_i - iT_0) \tag{7-34}$$

$$VS_{sinc}^{lc}{}^{W_m}(f_i) = \sum_{i=0}^{F_{W_m}-1} N_{lc}{}^{W_m}[j]\, sinc(\square_i - iT_0) \tag{7-35}$$

The visual speech signals driven by concatenating the normalized ratio of upper to corner and lower to corner visual speech sample sets is:

$$VS_{sinc}^{uclc^{W_m}}(f_i) = \sum_{i=0}^{2F_{W_m}-1} uclc^{W_m}[n]\, sinc(f_i - iT_0) \tag{7-36}$$

where $T_0 = 1$. The visual speech sample sets are also constructed by sinc function as the basic functions. The constructed visual speech signals are compared with the ones that are obtained by the BLI.

- **UPPER VISUAL SPEECH SIGNALS**

The visual speech signals ${VS_{sinc}^{u}}^{W_m}(f_i)$, $m = \{1,2,3,\dots,13\}$ are shown in Figure 7-52.

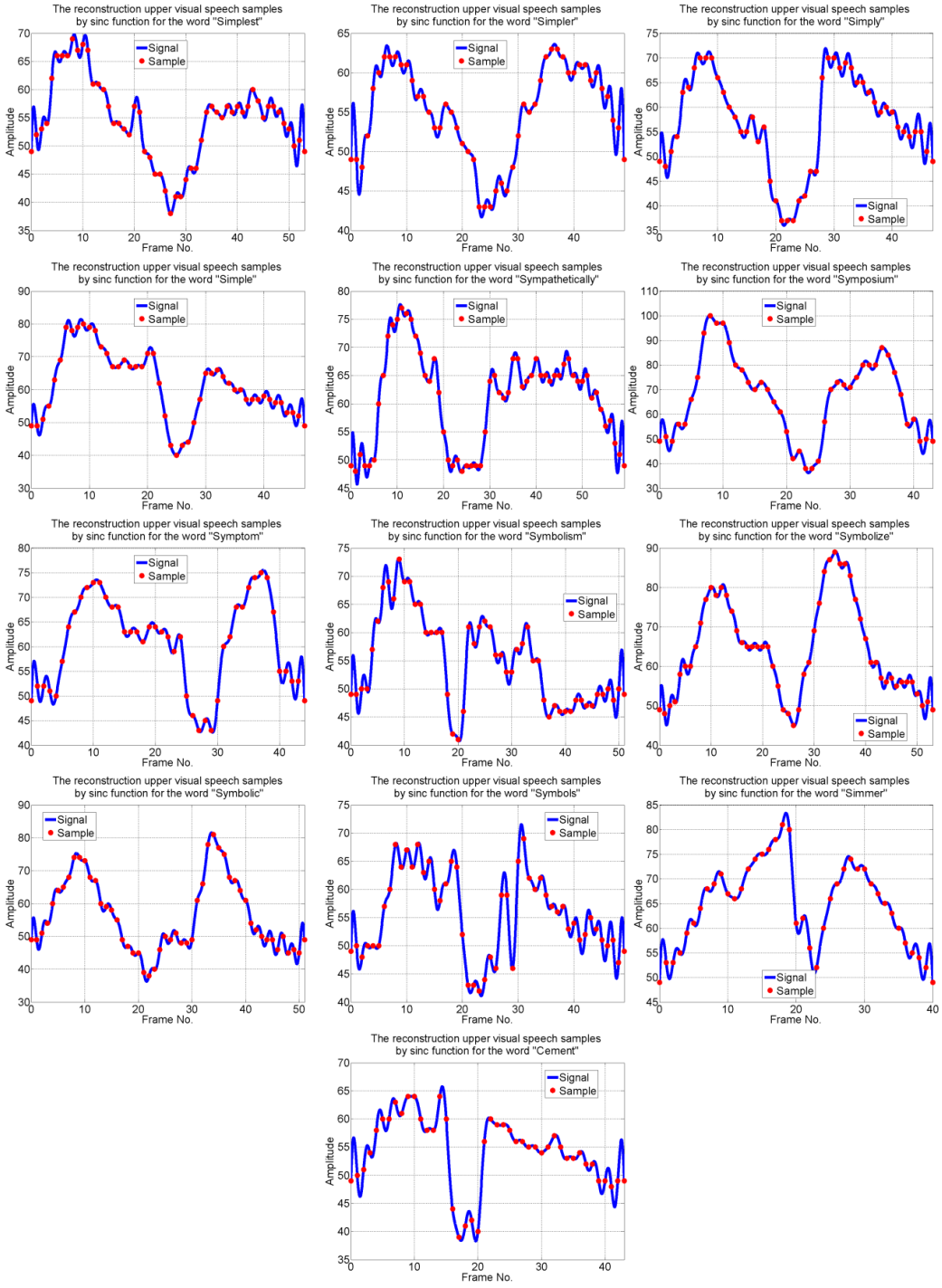


Figure 7-52: The upper visual speech signals of the selected words constructed by the sinc function

- **LOWER VISUAL SPEECH SIGNALS**

The visual speech signals $VS_{sinc}^{l}{}^{W_m}(f_i), m = \{1,2,3,\dots,13\}$ are shown in Figure 7-53.

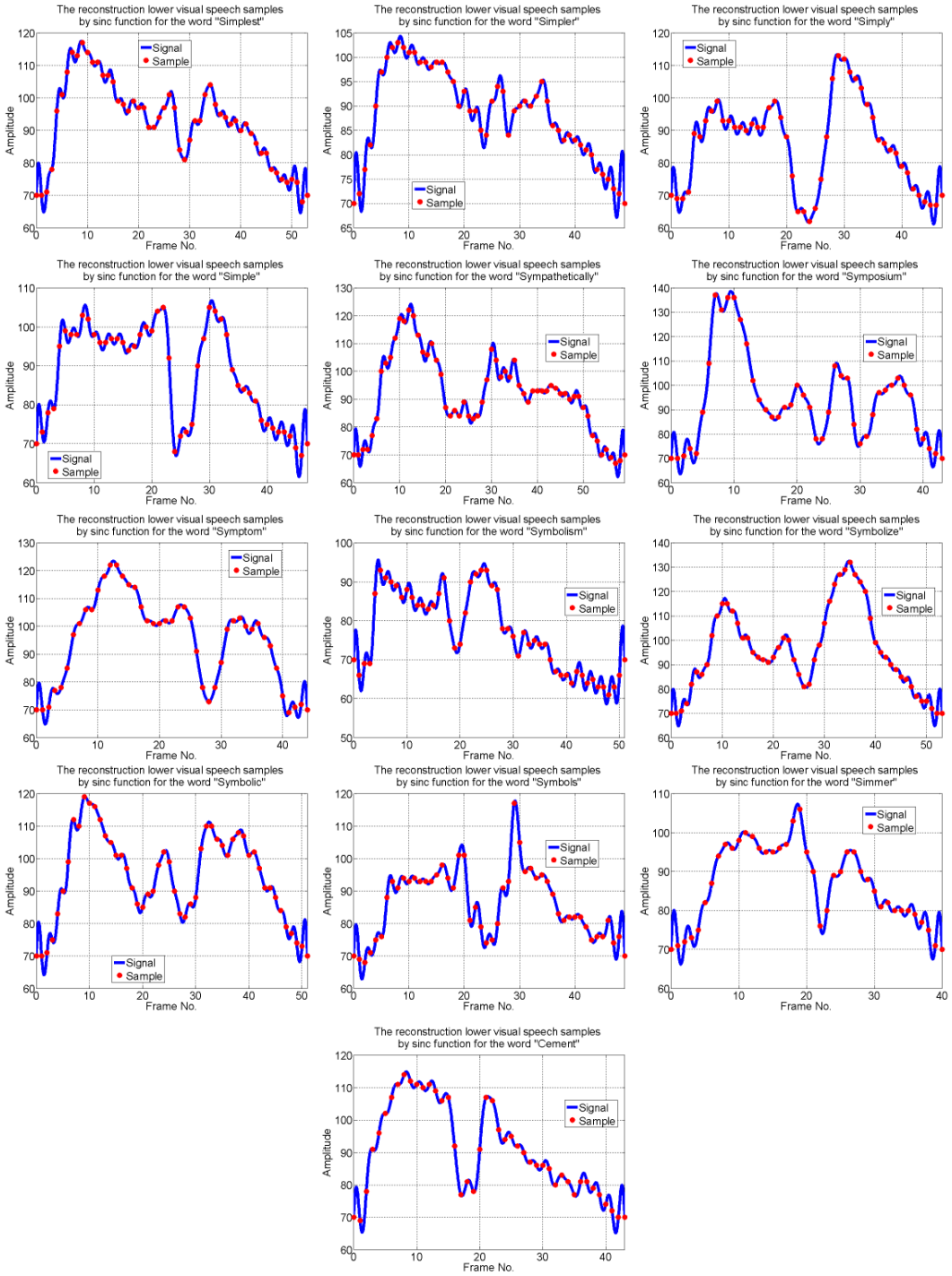


Figure 7-53: The lower visual speech signals of the selected words constructed by the sinc function

- **CORNER VISUAL SPEECH SIGNALS**

The visual speech signals $VS_{sinc}^{c}{}^{W_m}(f_i), m = \{1,2,3,\dots,13\}$ are shown in Figure 7-54.

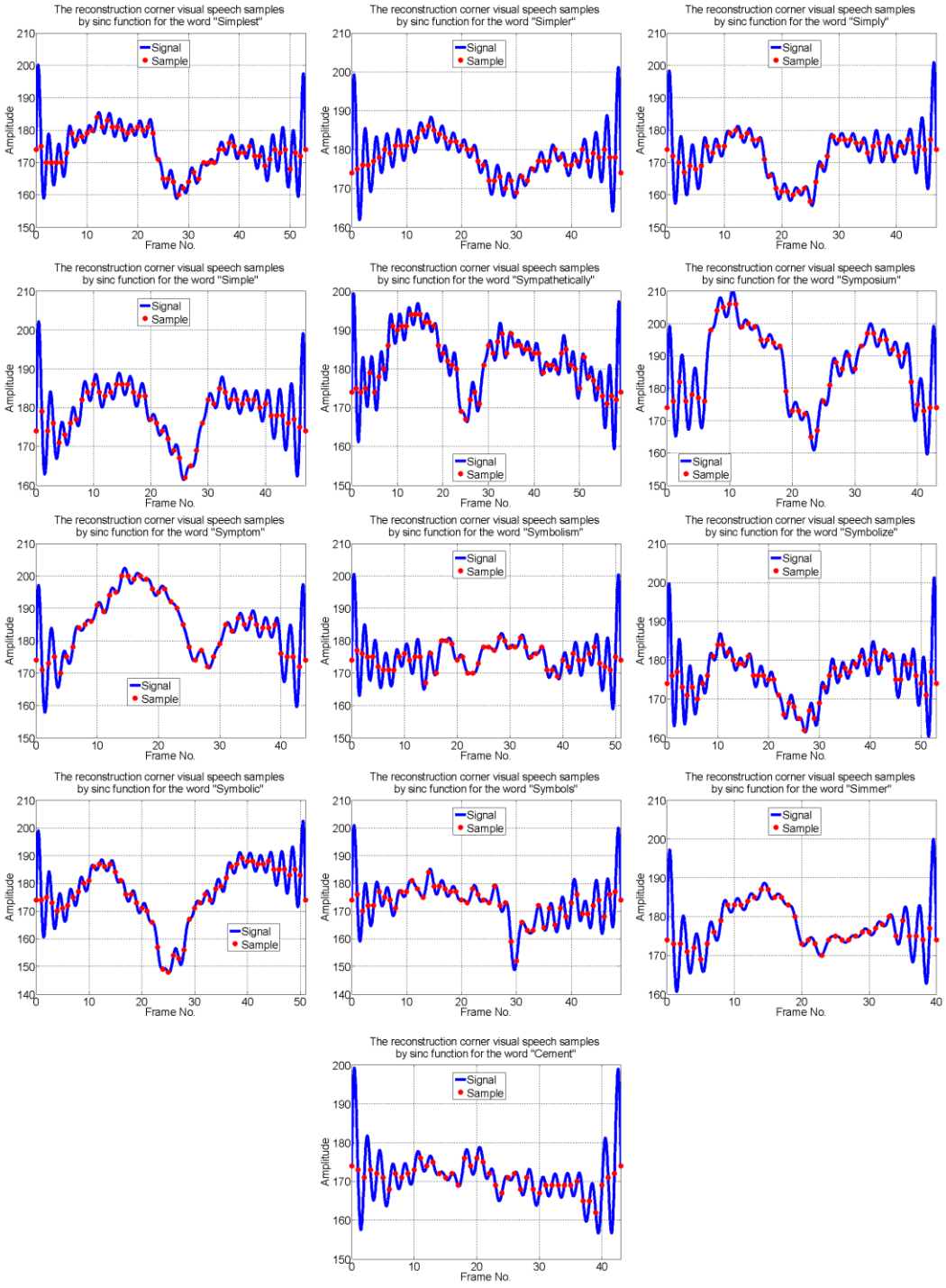


Figure 7-54: The corner visual speech signals of the selected words constructed by the sinc function

- **VISUAL SPEECH SIGNALS $VS_{sinc}^{uc}{}^{W_m}(f_i)$ AND $VS_{sinc}^{lc}{}^{W_m}(f_i)$**

In this section, the visual speech signals are constructed from $N_{uc}{}^{W_m}[j]$ and $N_{lc}{}^{W_m}[j], m = \{1,2,3,\dots,13\}$ which are denoted by $VS_{sinc}^{uc}{}^{W_m}(f_i)$ and $VS_{s\square nc}^{lc}{}^{W_m}(f_i), m = \{1,2,3,\dots,13\}$. The visual speech signals $VS_{sinc}^{uc}{}^{W_m}(f_i)$, are shown in Figure 7-55.

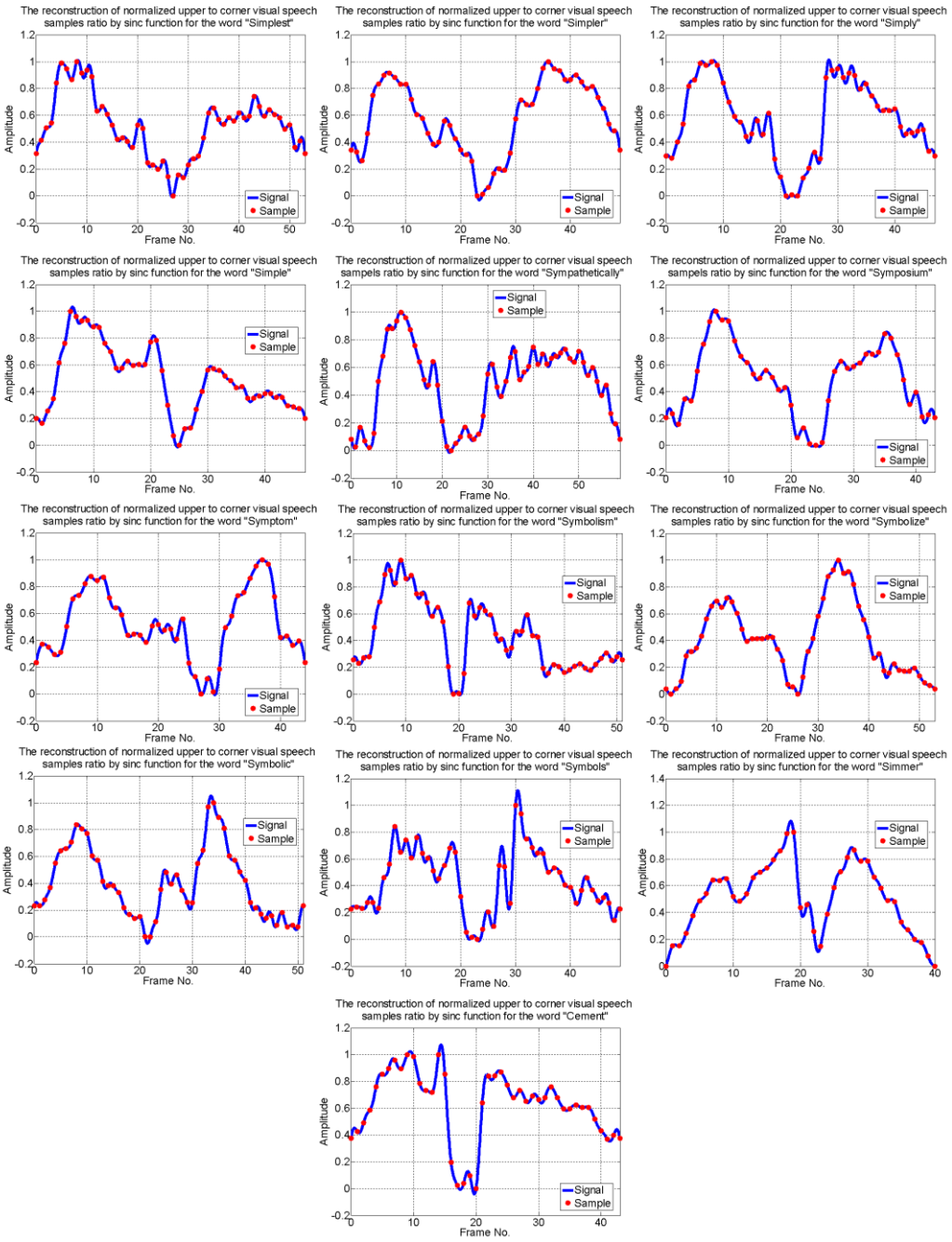


Figure 7-55: The visual speech signals constructed by the normalized ratio of upper to corner visual speech sample sets for all selected words constructed by the sinc function

The visual speech signals $VS_{sinc}^{lc}{}^{m}(f_i)$, $m = \{1,2,3,\ldots,13\}$ are shown in Figure 7-56.

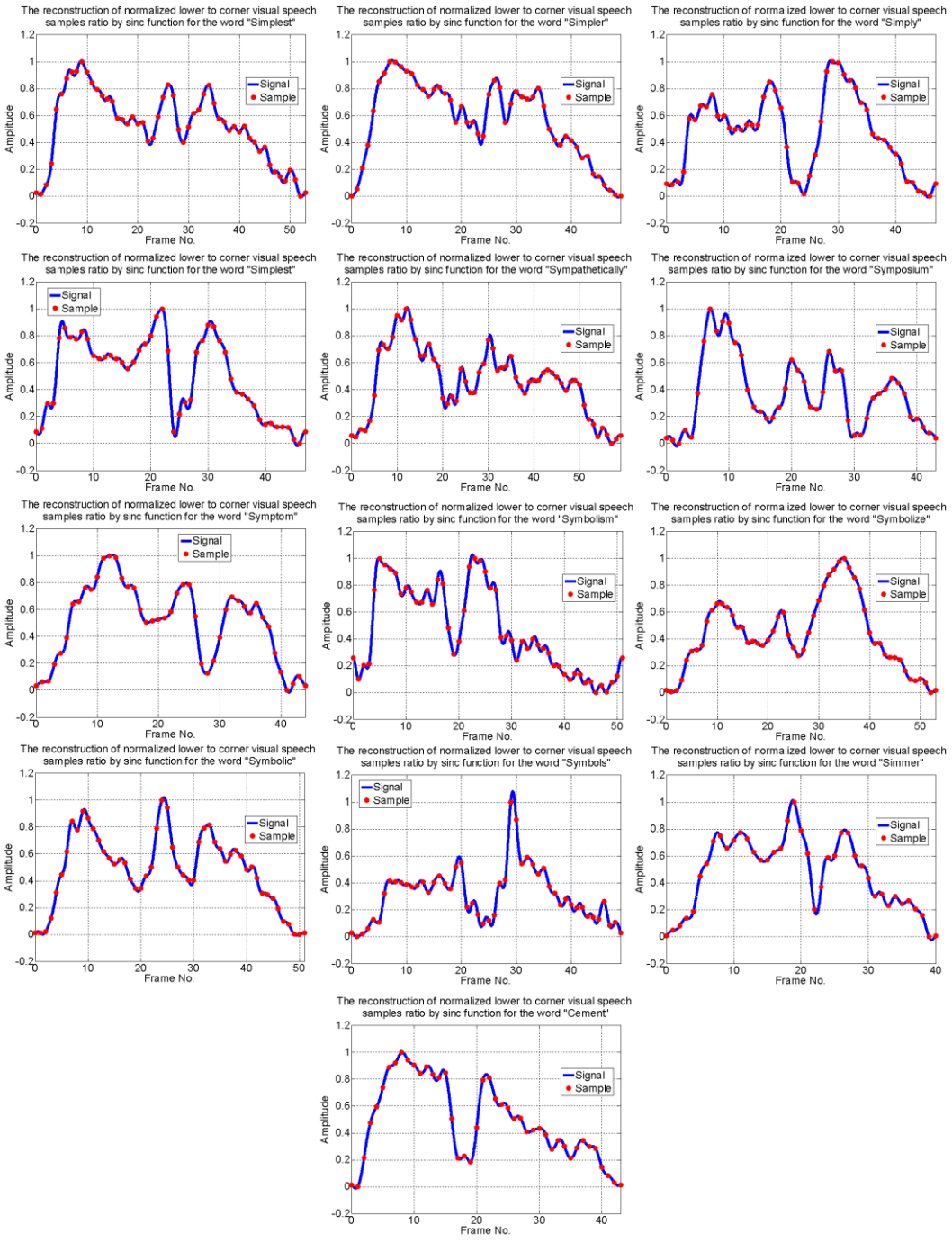


Figure 7-56: The visual speech signals constructed by the normalized ratio of lower to corner visual speech sample sets for all selected words constructed by the sinc function

- **CONCATENATED VISUAL SPEECH SIGNAL**

In this section, the visual speech signals are constructed from $uclc^{W_m}[n]$, $m = \{1,2,3,\dots,13\}$ where denoted by $VS_{sinc}^{uclc^{W_m}}(f_i)$, $m = \{1,2,3,\dots,13\}$ and shown in Figure 7-57.

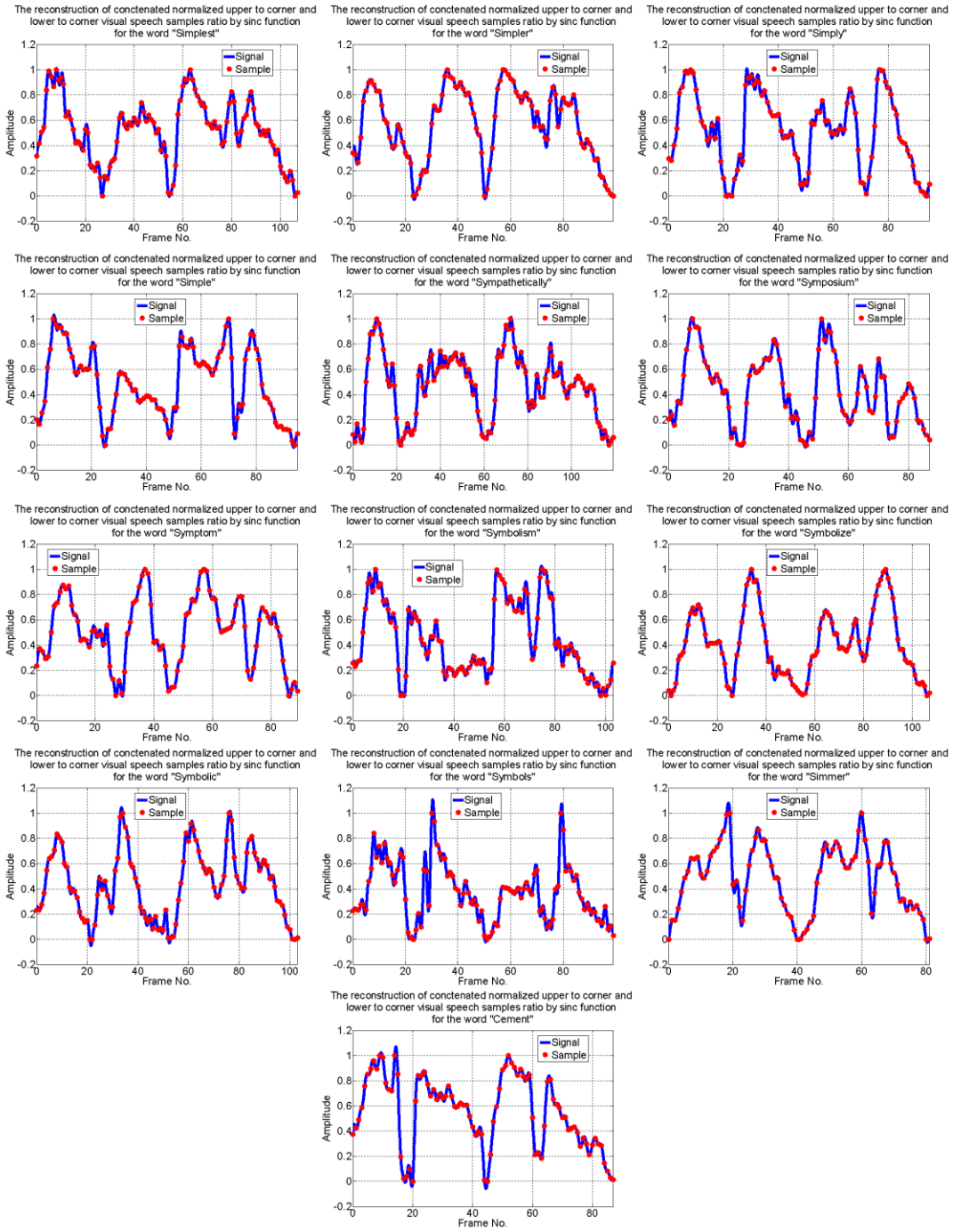


Figure 7-57: The concatenated visual speech signals constructed by the normalized ratio of upper to corner and lower to corner visual speech sample sets for all selected words constructed by sinc function

The constructed visual speech signals with sinc function in comparison to the BLI visual speech signals (Figure 7-3 to Figure 7-5), exhibit more curve fluctuations especially in the beginning and ending of signals. This is due to defining sinc function as the basic function for constructing the signals. Sinc function has better performance for fast sample sets while the slower sample set (Figure 7-55) are not accurately follow the amplitude trends.

- **METHODS FOR RESULTS EVALUATION**

Evaluating the results has been performed by comparing the visual speech signals ${VS_{BLI}^{u}}^{W_m}(f_i)$, ${VS_{BLI}^{l}}^{W_m}(f_i)$, ${VS_{BLI}^{c}}^{W_m}(f_i)$, ${VS_{BLI}^{uc}}^{W_m}(f_i)$, ${VS_{BLI}^{lc}}^{W_m}(f_i)$ and $VS_{BLI}^{uclc^{W_m}}(f_i)$, $m = \{1,2,3,\dots,13\}$ with the visual speech signals ${VS_{sinc}^{u}}^{W_m}(f_i)$, ${VS_{sinc}^{l}}^{W_m}(f_i)$, ${VS_{sinc}^{c}}^{W_m}(f_i)$, ${VS_{sinc}^{uc}}^{W_m}(f_i)$, ${VS_{sinc}^{lc}}^{W_m}(f_i)$ and $VS_{sinc}^{uclc^{W_m}}(f_i)$, $m = \{1,2,3,\dots,13\}$. The schematic of process is shown in Figure 7-58 where for the simplicity the visual speech signals constructed by the sinc function (ideal signal construction) are noted by $I(.)$.

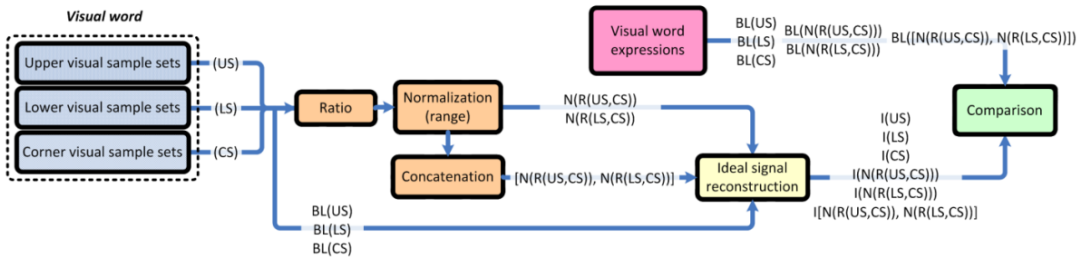


Figure 7-58: Comparing the visual speech signals and visual speech

The sinc function will be evaluated and compared to the corresponding BLI results.

- **COMPARISON OF ERROR BETWEEN THE VISUAL SPEECH SIGNALS CONSTRUCTED BY THE BLI AND THE SINC FUNCTION**

In this section, the RMSE of the constructed visual speech signals by the BLI and sinc function from the visual speech sample sets $u^{W_m}[n]$, $l^{W_m}[n]$ and $c^{W_m}[n]$, ${N_{uc}}^{W_m}[n]$, ${N_{lc}}^{W_m}[n]$ and $uclc^{W_m}[n], m = \{1,2,3,\dots,13\}$ and $n = \{0,1,2,\dots,F_{W_m} - 1\}$ are calculated. The RMSE is calculated in MATLAB environment by:

$$RMSE = \frac{norm(abs(S1 - S2))}{sqrt(size(S1, S2))} \quad (7\text{-}37)$$

- **VISUAL SPEECH SIGNALS** $\boldsymbol{VS_{BLI}^{u}}^{W_m}(\boldsymbol{f_i})$, $\boldsymbol{VS_{BLI}^{l}}^{W_m}(\boldsymbol{f_i})$ **RMSE VERSUS** $\boldsymbol{VS_{BLI}^{c}}^{W_m}(\boldsymbol{f_i}), \boldsymbol{m} = \{\boldsymbol{1}, \boldsymbol{2}, \boldsymbol{3}, \dots, \boldsymbol{13}\}$ AND $\boldsymbol{VS_{sinc}^{u}}^{W_m}(\boldsymbol{f_i})$, $\boldsymbol{VS_{sinc}^{l}}^{W_m}(\boldsymbol{f_i})$ AND $\boldsymbol{VS_{sinc}^{c}}^{W_m}(\boldsymbol{f_i})$

The RMSE of constructed visual speech signals ${VS_{BLI}^{u}}^{W_m}(f_i)$, ${VS_{BLI}^{l}}^{W_m}(f_i)$ and ${VS_{BLI}^{c}}^{W_m}(f_i), {VS_{sinc}^{u}}^{W_m}(f_i)$, ${VS_{sinc}^{l}}^{W_m}(f_i)$ and ${VS_{sinc}^{c}}^{W_m}(f_i), m = \{1,2,3,\dots,13\}$ are shown in Figure 7-59.

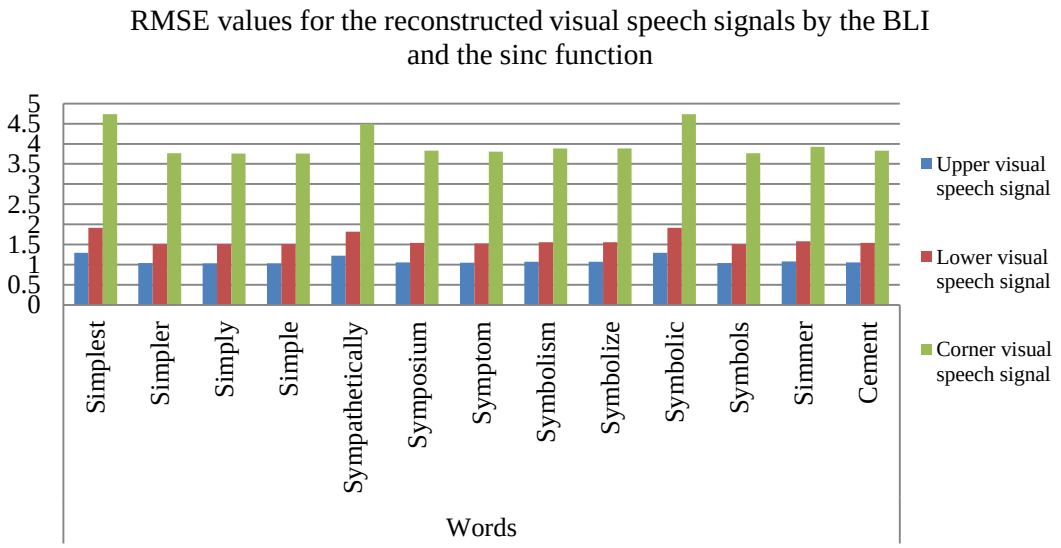


Figure 7-59: The RMSE values of the upper, lower, and corner visual speech signals constructed by the sinc function

The most difference between the visual speech signals appears in the corner visual speech signals.

- **VISUAL SPEECH SIGNALS** $\boldsymbol{VS_{BLI}^{uc}}^{W_m}(\boldsymbol{f_i})$ **AND** $\boldsymbol{VS_{BLI}^{lc}}^{W_m}(\boldsymbol{f_i})$ **VERSUSN** $\boldsymbol{VS_{sinc}^{uc}}^{W_m}(\boldsymbol{f_i})$ **AND** $\boldsymbol{VS_{sinc}^{lc}}^{W_m}(\boldsymbol{f_i})$

The RMSEs of constructed visual speech signals ${VS_{BLI}^{uc}}^{W_m}(f_i)$ and ${VS_{BLI}^{lc}}^{W_m}(f_i)$, $m = \{1,2,3,\dots,13\}$versus ${VS_{sinc}^{uc}}^{W_m}(f_i)$ and ${VS_{sinc}^{lc}}^{W_m}(f_i)$, $m = \{1,2,3,\dots,13\}$ are shown in Figure 7-60.

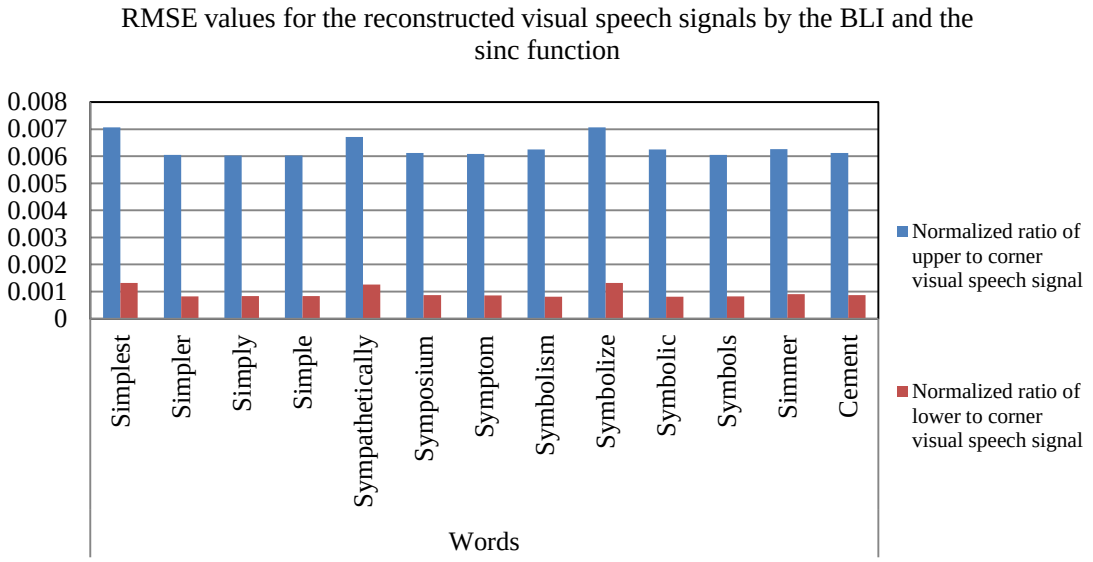


Figure 7-60: The RMSE values of the visual speech signals constructed by the sinc function from the normalized ratio of upper to corner and lower to corner visual speech sample sets

- **VISUAL SPEECH SIGNALS $VS_{BLI}^{uclc^{W_m}}(f_i)$ VERSUS $VS_{sinc}^{uclc^{W_m}}(f_i)$**

The RMSEs of constructed visual speech signals $VS_{BLI}^{uclc^{W_m}}(f_i)$ and $VS_{sinc}^{uclc^{W_m}}(f_i)$, $m = \{1,2,3,\dots,13\}$ are shown in Figure 7-61.

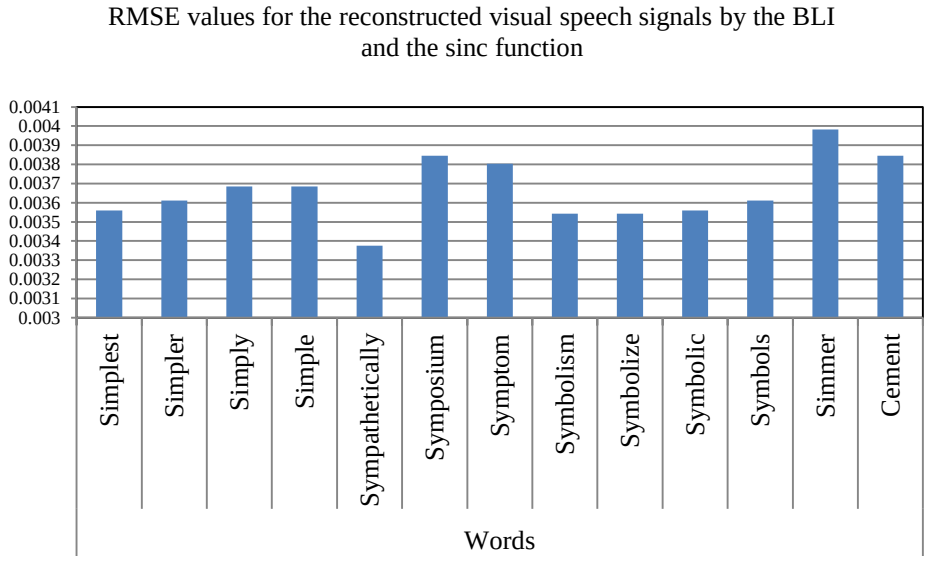


Figure 7-61: The RMSE values of the visual speech signals constructed by the sinc function from the concatenated normalized ratio of upper to corner and lower to corner visual speech sample sets

The maximum RMSE level of all types of visual speech signals shown Figure 7-59 to Figure 7-61 belongs to the constructed signal of the visual speech sample sets $c^{W_m}[n], m = \{1,2,3,\dots,13\}$. This effect can be observed by referring to Figure 7-5 and Figure 7-55 where the visual speech signals ${VS^c_{sinc}}^{W_m}(f_i),\ m = \{1,2,3,\dots,13\}$ are fluctuating near the boundaries of sample sets intervals. This effect also can be described due to the higher amplitudes of the visual speech sample sets and $c^{W_m}[n], m = \{1,2,3,\dots,13\}$ than $u^{W_m}[n]$ and $l^{W_m}[n]$. Therefore, the construction method using the sinc function is sensitive to the amplitudes of samples.

- **CORRELATION COEFFICIENTS**

The correlation coefficients of constructed visual speech signals ${VS^u_{BLI}}^{W_m}(f_i)$, ${VS^l_{BLI}}^{W_m}(f_i)$ and ${VS^c_{BLI}}^{W_m}(f_i) m = \{1,2,3,\dots,13\}$ and ${VS^u_{sinc}}^{W_m}(f_i)$, ${VS^l_{sinc}}^{W_m}(f_i)$ and ${VS^c_{sinc}}^{W_m}(f_i), m = \{1,2,3,\dots,13\}$ are shown in Figure 7-62.

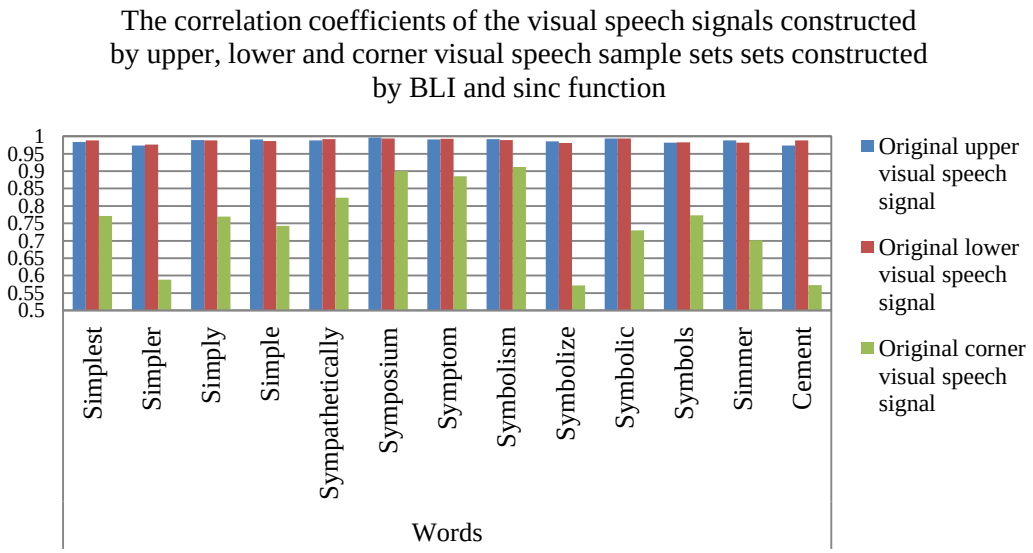


Figure 7-62: The correlation coefficients of the visual speech signals constructed by the BLI and the sinc function from the concatenated normalized ratio of upper to corner and lower to corner visual speech sample sets

The correlation coefficients of constructed visual speech signals ${VS^{uc}_{BLI}}^{W_m}(f_i)$ and ${VS^{lc}_{BLI}}^{W_m}(f_i)$, ${VS^{uc}_{sinc}}^{W_m}(f_i)$ and ${VS^{lc}_{sinc}}^{W_m}(f_i), m = \{1,2,3,\dots,13\}$ are shown in Figure 7-63.

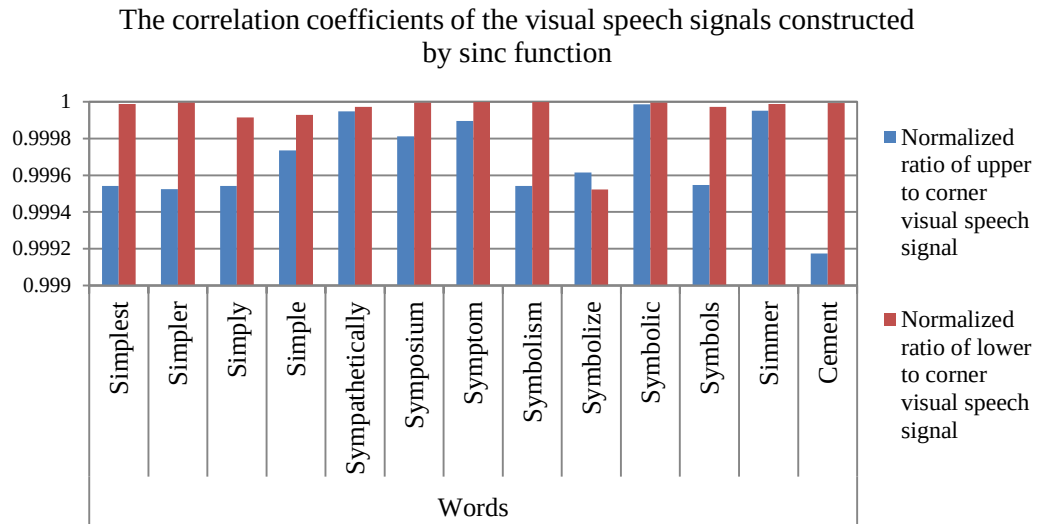


Figure 7-63: The correlation coefficients of the visual speech signals constructed by the sinc function from the normalized ratio of upper to corner and lower to corner visual speech sample sets

The RMSEs of constructed visual speech signals $VS_{BLI}^{uclc^{W_m}}(f_i)$ and $VS_{sinc}^{uclc^{W_m}}(f_i)$, $m = \{1,2,3,\dots,13\}$ have been shown in Figure 7-64.

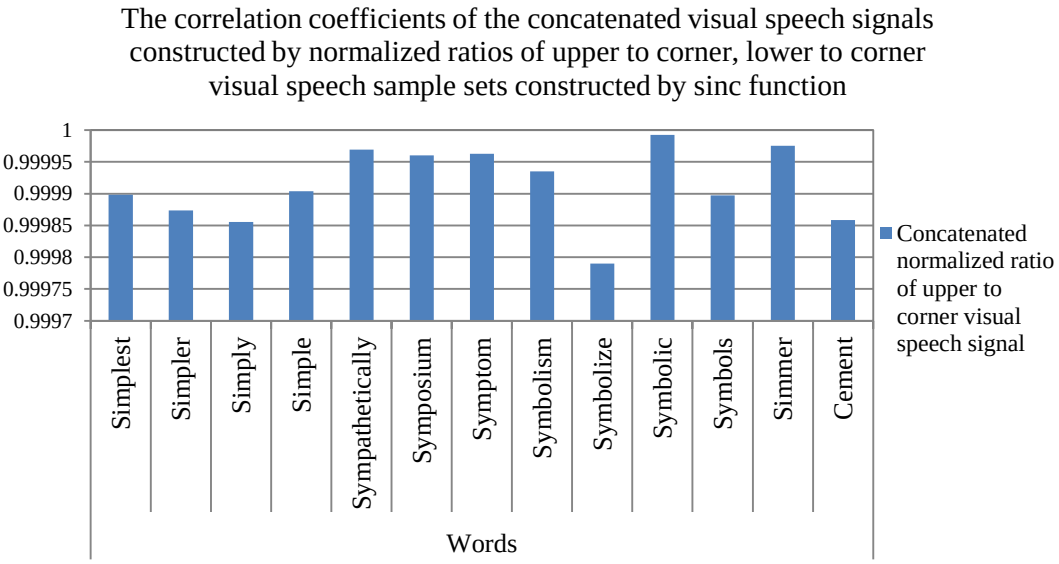


Figure 7-64: The correlation coefficients of the visual speech signals constructed by the sinc function from the concatenated normalized ratio of upper to corner and lower to corner visual speech sample sets

By comparing the correlation coefficients in Figure 7-62 to Figure 7-64, the least correlated visual speech signals belong to the constructing visual speech sample

sets $c^{W_m}[n]$, $m = \{1,2,3,\dots,13\}$, as it previously observed by the corresponding RMSE values.

7.8. SUMMARY

In this chapter, the core of visual speech signature allocation to the visual words had been introduced. The mathematical expressions of the visual speech data obtained from the upper, lower, and corner visual speech sample sets, called the visual speech signals. In order to validate the mathematical expressions, a geometrical method introduced for relating the sampling frequency to the samples positions. The graphical relations of the sampling frequency and errors are determined. Ideally, the error of mathematical expressions reduces by increasing the sampling frequency. However, since the number of decimal digits needs to be finite, the error will be fixed after a certain sampling frequency.

The normalized rational relations of visual speech sample set is chosen for providing more compact version of visual speech signal as well as preserving the scaling ability for the visual speech signals. The concatenated versions of the normalized ratios of the visual speech signals are driven numerically. In this stage, the concept of signatures for the visual words is introduced. The first type of allocated signature to the visual speech signals is chosen by the barcode identification. The second type of signatures is suggested as digital representations of the visual speech signals. This model consists of two components called signal quantization and the Huffman coding. In the third approach, the visual speech signals are represented in the 2D and 3D Cartesian space. These graphical approaches, which is highly related as what human considers as signature, concluded the visual speech signatures of visual words. In this stage, the idea of relating the upper, lower, and corner visual speech signals via a generic model of lip is introduced. Using the Bezier closed curve for relating the visual features as a closed curve is the motivation for creating a generic lip model. It is observed that the Bezier closed curve can be modelled by an ellipse. Three different ellipses are defined that included the three visual feature points. These ellipses are averaged for representing the lip with a template of a single ellipse. By allocating the sequence of such ellipse to visual speech samples, which are also varying according to the visual speech signals, a volumetric representation of articulating lip is defined. The averaged ellipse is included the visual speech sample sets and defined a form of

lip template simultaneously. In Section 7.6, the visual speech sample sets are also constructed by the discrete-to-continuous-time signal conversion method. More specifically, the sample sets are constructed via an interpolation that used sinc function as the basic function. Finally, the comparisons in terms of RMSE and correlation coefficients between the visual speech signals constructed using the BLI and the corresponding visual speech signal constructed using sinc function are performed.

Chapter 8

CONCLUSION AND RECOMMENDATIONS FOR FUTURE WORK

8.1. CONCLUSION

In this chapter, the overall achievements of the work reported in the thesis are presented. The concept of viseme does not provide a global and unique definition in terms of the visual representations of phonemes. Furthermore, the visemes have been defined as separated segments, which have been concatenated with weighing function to construct visual phoneme representations. In terms of related transcriptions, the visual phonemes did not analyse and arrange systematically. In this thesis, in order to overcome these issues, the visual speech clues have been expressed by the mathematical equations. Furthermore, the visual clues have been categorized based on the mutual phonemic structures in a set of words. In order to obtain such relation, a method of designing the transcribed words (corpus) has been introduced. The designing procedure has been based on hierarchical analysis of phoneme sequences. More specifically, a set of words are chosen from the TIMIT's vocabulary dictionary with maximum appearance according to the statistical analysis of three sequences of consonants and vowel (triphone) as /CVC.../. The number of selected words is 13. In this manner, the visual similarities corresponding to the first three phonemes have been also preserved. Furthermore, the Co-articulation effect appears in the visual domain that eliminates the usage of other components to induce the Co-articulation effect.

The visual information of the articulating lips obtained by choosing a region of interest (which is a rectangle containing the lip) which contains three feature points located on the outer upper, lower and corner lip contours. The visual feature points

have been represented by the pixel values on the video frames. Speakers articulated the selected words in the presence of the words audio file (for increasing better articulation). The frequency of recording set to 30 frames per second. Afterward, the video files processed and visual features have been extracted based on approximating the words frame size from the audio file with anticipation effect. The process of visual data extraction has been performed manually. The sequence of extracted pixel values for each word has been called visual speech sample sets. Since these guidelines did not match to the extracted data, the statistical trimmed mean in different versions of the extracted data has been chosen to find the appropriate frame lengths of articulated words. In comparison to the automatic colour-based methods of visual feature acquisition, the manual extraction is superior in term of accuracy (see Section 4.5) although the process of extraction is time consuming. The algorithms employed in colour-based analysis are under the effect of background, skin colours, and environmental lightening. These factors tend to reduce the accuracy of extracted data. The number of speakers is also the subject of argument since it can vary widely. Since the possibility of obtaining accuracy in collecting the data samples has been not the objective of this work, the visual data has been extracted from two non-native English speakers. Undoubtedly, increasing the number of speakers will increase the accuracy of visual data and mathematical expressions. The number of speakers in this work has been two.

The extracted visual speech sample sets have been suffering from two issues. The first issue has been the misbalancing between starting and stopping samples for each set and second has been the amplitude differences in the starting samples. The common roots of these issues have been the unintentional speakers' head movements, imperfect articulation as well as non-ideal manual extractions. The first issue has been addressed by subtracting a line connecting the sample set's end points for each sample set and the second issue has been solved by adding (biasing) all the balance visual speech sample sets by the trimmed mean obtained from all the first samples in corresponding sets. The statistical characteristics of the visual speech sample sets have been examined to obtain more information about their families of distributions they belong. By examining the distribution of visual speech sample sets, it has been observed that they do not match to the common distributions. Therefore, the Pearson system has been used for determining of the distribution family type of the visual

speech sample sets. By applying the visual speech sample sets to the Pearson system before endpoints adjustment and biasing, the density function for the upper visual speech sample sets have been determined as unknown while the lower and corner visual speech sample sets have been detected as beta distribution. After correcting the misbalancing and adding biasing to the visual speech sample sets, the Pearson system detected the upper and lower visual speech sample sets belonging to beta distribution family while the corner visual speech sample sets' distribution type has been detected as unknown. In that part, a GUI of lip, which has been used for observing the visual speech sample sets, has been generated to exhibit an animation of the processed data.

The processed visual data has been inserted to the next stage for mathematical modelling. The mathematical expressions for the visual speech sample sets have been driven by an interpolation method called the Barycentric Lagrange Interpolation (BLI). Each word has been expressed with three visual speech signals, called visual word.

In the beginning, the BLI reconstruction performance has been evaluated by examining the errors appearing in reconstruction of two sets of deterministic functions. The error has been calculated from two scenarios. In the first scenario, the length of provided samples to the BLI has been varying while their amplitudes kept fixed. In the second scenario, the length of provided sample sets to the BLI is fixed and the amplitudes varied. In the first scenario, the error of interpolation reduces while in the second scenario the error of interpolation increased as the amplitudes of samples increased. The next important characteristic of the BLI method has been related to the expression that it defines. One of the main aims of formulating the visual speech sample sets has been to recover them back from the mathematical formula. In Figure 7-16 to Figure 7-20 the validation for extracting the visual speech sample sets from the mathematical expression in terms of number of sampling rates and corresponding errors have been represented. It can be observed that there has been a common characteristic in all graphs. There has been a region, where error reaches to a minimum, have been increasing the sampling rate do not affected it. In other words, the mathematical expressions of visual speech signals and their concatenated versions represent the visual words with errors either graphical or in case of recovering the visual speech sample sets that could not be improved via

increasing the sampling rate. According to Figure 7-21 and Figure 7-22, the minimum error in extracting the visual speech sample sets from the expressions has been related to the sampling rate. Expect the words 'Simply' and 'Symbolize', the corner visual speech signals had higher error during recovering the sample sets. Similarly, the upper visual speech signals had lowest error rates. Interestingly, the lower visual speech signal for the word 'Symptom' has slightly higher error than 0.004 with sampling frequency less than 100. On the other hand, all three visual speech signals in words 'Simply', 'Simple' and 'Sympathetically' must have generated by minimum 1000 samples to obtain there corresponding error rates. In conclusion, the maximum error for recovering the samples from mathematical expressions is 0.01018 while the minimum achievable error is 0.0005845 for 1000 samples. The suggested method uses the geometrical relations in recovering the samples with a minimum error. The error has been calculated if the sampling rate increased and the error converged to a constant value while the number of decimation in the formula is fixed.

The visual speech sample sets have been related to each other in order to reduce the number of visual speech signals corresponding to each visual representation of words. To do such reduction, the ratios of the upper to corner and lower to corner sample sets have been calculated and then normalized. The normalization process caused the amplitude unification of the ratios of sample sets. Further reduction of the number of samples set to one has been achieved by concatenating the normalized ratios of the upper to corner and lower to corner visual speech sample sets. These visual speech signals have been called the visual word's signatures. In this stage, the mathematical expressions have been provided by BLI except the concatenated versions of visual speech sample sets due to software limitation in calculating the results. Nevertheless, if the software has been able to find the expressions, still it would be highly impractical for using the mathematical expressions since they would be extremely long. Therefore, the numerical values of the signal have been suggested.

The concatenated normalized ratios of the upper to corner and lower to corner visual speech sample sets have been defined as the numerical signatures of visual words. The concept of signature has been introduced as any unique form of visual word representation. In order to obtain more simplified expressions and identification

purposes, the visual words with different descriptions have been associated to signature representations with four methods. Therefore, the visual words have been characterized by allocating barcode, digital representations, and scattering plots to the consisting visual speech signals as well as volumetric representations of each visual word. These four versions of visual words could be used in the visual front-end of AVASR as well as animation systems. The barcode and digital representations of such visual speech signals also uniquely express the visual words. The barcodes duty cycles and number of bars have been characterized the visual words numerically. The digital representation of the visual words obtained by quantization of the visual speech signals which has been followed by the Huffman coding scheme. In case of scattering plot, the 2D/3D representations of the visual speech signals, which has been driven from the normalized ratios of upper to corner and lower to corner visual speech sample sets, represent signatures similar to those used in daily. Such type of representation also allocates a unique graphical signature to each visual word. The forth type expressed the visual words by a volumetric model related to the lip dimension. The idea of volumetric representation of visual words originates from modelling the pairs of visual features by Bezier closed curves. The resulted curves represent ellipsoid characteristics that have been used for final relations of the feature points. The upper and corner visual features, the lower and corner visual features and the upper and lower visual features have been included in ellipse equation. In this part, the three BLI expressions for a visual word can be substituted in the ellipsoids equations. This leads to relate the BLI expression to gather. By averaging these three ellipses, a single ellipse has been defined that models the lip's geometry. The sequences of the ellipses, which have been calculated by the visual speech sample sets, represent the volumetric representation of a visual word aligned on the frame numbers as the third dimension.

Beside the BLI of lip's movements, the visual sample sets have been constructed by ideal signal reconstruction method used in the field of signal processing theory. The core of the reconstruction method consists of using the sinc function as the basis function. These two sets of visual words have been compared in terms of RMSEs and correlation coefficients. The corner visual speech signals and the BLI of the ratios of the upper to corner visual sample sets have been shown maximum RMSE. Similarly,

the correlation coefficients of these two sets of visual speech signals have been the lowest among the others.

The restrictions of processing the visual speech signals in order to store in the visual word database have to be considered. The main limitation belongs for driving the explicit mathematical expressions when the normalized rational upper to corner and lower to corner visual speech sample sets are concatenating. In the MATLAB environment using the 'Symbolic Math Toolbox', it has been not possible to display lengthy expressions. More specifically, it has been not practical for the software to display more than 25,000 characters in its output display. Therefore, the visual words mathematical expressions have been defined for upper, lower, and corner visual speech sample sets and the normalized ratios of upper to corner and lower to corner visual speech sample sets. The visual speech signal of concatenating of the normalized rational upper to corner and lower to corner visual speech sample sets has been stored only numerically.

The next important parameter is the computational cost of the proposed approaches. The workstation used for simulation had the system properties of Intel (R) Core (TM) 2 Duo CPU E8300 with the frequency of 2.83GB for each core and 3.48GB available RAM. The total time consumption for processing the outcomes of this work is 15 minutes and 48 seconds, which are represented in details in Table 8-1. These values have been associated by 'tic' command in the MATLAB environment. The most time consuming process belongs to the mathematical formulation of visual speech signals.

Table 8-1: The processing time consumption

Time consumption (second)	Method
0.990	Visual speech signal generation
6.905	Barcode
2.499	Quantized signals
0.295	Digital representation
28.876	2D/3D representation
3.694	Volumetric representations
66.428	Evaluation
842.743	Mathematical formulation
952.430	Total

As it is represented in Table 7-16, there are six groups of data sets. They have been associated with mathematical expressions, graphical representation, barcode, digital

representation, and 2D/3D signatures and volumetric visual words. According to the criteria and the software computational limitations, the availability of each entity is marked. These expressions are stored under the name of visual-based vocabulary database. Furthermore, in Table 8-2 after applying the BLI and ideal construction to the data sets, the mathematical expressions, barcode, and their digital representations have been allocated by the specific notation. If the visual data has mathematical expressions, graphical representation, barcode, digital representation, 2D signature, or 3D signature, it will be noted by Tick (√) sign, otherwise by Cross sign(×). The evaluations of visual speech sample sets recovery and BLI method in term of error analysis has been listed in Table 8-2.

Table 8-2: The list of structure and type of outcomes in order to store in the visual signature-based vocabulary database

		Mathematical expressions	Graphical representation	Barcode	Digital representation	2D signature	3D signature
Visual data	Upper visual speech sample set	✓	✓	×	×	✓	×
	Lower visual speech sample set	✓	✓	×	×	✓	
	Corner visual speech sample set	✓	✓	×	×	✓	
	Normalized ratio of upper to corner visual speech sample	✓	✓	✓	✓	✓	×
	Normalized ratio of lower to corner visual speech sample	✓	✓	✓	✓		×
	Concatenation of the normalized ratio of upper to corner and lower to corner visual speech sample	×	✓	✓	✓	✓	×

Table 8-3: List of mathematical notations of components in the visual signature-based vocabulary database

			Mathematical expressions	Barcode			Quantized visual speech signals	Digital representation
				PWM expression	BDC	No. of bars		
Visual word	BLI	Upper visual speech signal	$VS_{BLI}^{u\ W_m}(f_i)$	-	-	-	-	-
		Lower visual speech signal	$VS_{BLI}^{l\ W_m}(f_i)$	-	-	-	-	-
		Corner visual speech signal	$VS_{BLI}^{c\ W_m}(f_i)$	-	-	-	-	-
		Normalised ratio of Upper to corner Visual speech signals	$VS_{BLI}^{uc\ W_m}(f_i)$	$BVS_{BLI}^{uc\ W_m}(f_i)$	NRUCBDC	NRUCB	$VS_{BLI}^{uc\ W_m}[f_i]$	$DNRVS_{BLI}^{uc\ W_m}$
		Normalised Ratio of lower to corner Visual speech signal	$VS_{BLI}^{lc\ W_m}(f_i)$	$BVS_{B\square I}^{lc\ W_m}(f_i)$	NRLCBDC	NRLCB	$VS_{BLI}^{lc\ W_m}[f_i]$	$DNRVS_{BLI}^{lc\ W_m}$
		Concatenation of the normalized ratio of upper to corner and lower to corner visual speech signals	$VS_{BLI}^{uclc W_m}(f_\square)$	$BVS_{BLI}^{uclc W_m}(f_i)$	ULCBDC	NCBDC	$VS_{BLI}^{uclc W_m}[f_i]$	$DRNVSC_{BLI}^{lc\ W_m}$
	Ideal construction by sinc function	Upper visual speech signal	$VS_{sinc}^{u\ W_m}(f_i)$	-	-	-	-	-
		Lower visual speech signal	$VS_{sinc}^{l\ W_m}(f_i)$	-	-	-	-	-
		Corner visual speech signal	$VS_{sinc}^{c\ W_m}(f_i)$	-	-	-	-	-
		Ratio of Upper to corner Visual speech signals	$VS_{sinc}^{uc\ W_m}(f_i)$	-	-	-	-	-
		Ratio of lower to corner Visual speech signal	$VS_{sinc}^{lc\ W_m}(f_i)$	-	-	-	-	-
		Concatenation of the normalized ratio of upper to corner and lower to corner visual speech signals	$VS_{sinc}^{uclc W_m}(f_i)$	-	-	-	-	-

Table 8-4: List of error evaluation regarding to processing the visual speech sample sets

		Error evaluation				
		Visual samples endpoints adjustment	The BLI evaluation	Recovering the visual speech sample sets	Quantization	Constructed visual speech signals from BLI vs. sinc function
Visual speech signals	$VS_{BLI}^{u}{}^{W_m}(f_i)$ $VS_{BLI}^{l}{}^{W_m}(f_i)$ $VS_{BLI}^{c}{}^{W_m}(f_i)$	✓	×	✓	✓	✓
	$VS_{BLI}^{uc}{}^{W_m}(f_i)$ $VS_{BLI}^{lc}{}^{W_m}(f_i)$	×	×	✓	✓	✓
	$VS_{BLI}^{uclc}{}^{W_m}(f_i)$	×	×	✓	✓	✓
Gaussian function and approximate polynomial from visual speech sample sets	Variable length and fixed amplitude sample sets	×	✓	×	×	×
	Fixed length and variable amplitude sample sets	×	✓	×	×	×

8.2. FUTURE WORK RECOMMENDATION

The suggested method for analysing the lip's movement provides a template for formulating the mathematical framework in representation of the categorized words. It is recommended to implement the visual words mathematical signatures and its variation in lip reading systems. In this case, a spoken word would be categorized to predict the next phoneme leading to narrowing down the search.

The usage of suggested methods also appears in visual speech synthesisers. The mathematical representation of visual words includes the articulation effect and provides lower complexity of the system. It leads to less computational cost. More

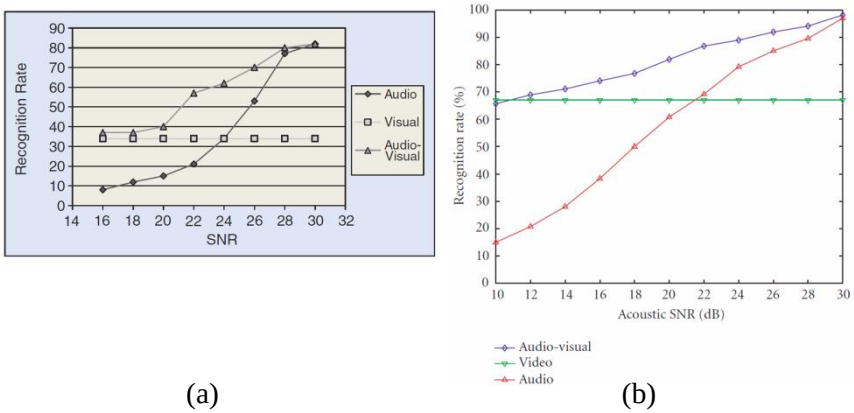
specifically, the author of the following possibilities of extending the proposed methods of this work as:

a) Categorizing an entire transcribed dictionary as its methodology suggested in Chapter 3.
b) Studying the transition between words in meaningful sentences.
c) Increasing the number of speakers of the target language that are sampled.
d) Studying the speed of lip's movements
e) Finding the relation of the visual words and the corresponding audible data.
f) Employing appropriate and practical components for visual feature detection.
g) Automating the process of visual word recognition (lip-reading).
h) Synthesizing the visual words (animation).
i) Defining the real-time implementation for the items 8 and 7.

APPENDIX

A VISUAL SPEECH SYNTHESIS

The visual speech synthesis or visual speech animation is the next category of human machine interaction that provides a user interface by modelling human facial expressions. The application of the interface is enhancing the perception of facial information for hearing impaired people. It can also be considered as complementary part in the AVSR systems in the visual front-end module. For example, enhancing the performance of a speech recognition system studied by (Chen, 2001) in presence of acoustic white noise that caused to variation of signal-to-noise ratio $SNR \in [16^{db}: 30^{db}]$ as it shown in Figure 8-1 (a) (Chen, 2001), while the recognition rate in presences of white Gaussian is obtained by (Nefian, et al., 2002) for $SNR \in [10^{db}: 30^{db}]$ in Figure 8-1 (b) (Nefian, et al., 2002).



(a) (b)

Figure 8-1: The recognition ratio versus SNR of the acoustic signal (Chen, 2001) (a) and comparison of the recognition rate of the audio-only, video only and CHMM-based audio-visual speech recognition (Nefian, et al., 2002) (b)

The era of facial animations started with the novel work of (Park, 1972), where he modelled a head with less than 100 polygons as in 3D. The animation is the result of analysing the key frames and a complex non-automated process. Two years later, he

developed a parametric model that is synchronized to speech, using linear interpolation of face geometry and parameters pertaining to facial expressions (Parke, 1974). In 1978, a model of facial expressions and measurement developed by (Ekman & Friesen, 1978) the facial action coding system (FACS) to measure and model facial expressions. Using FACS units and a muscle base model developed by (Waters, 1987) resulted in a realistic 3D face model. More sophisticated animating face, in which speech became synchronized to facial animations, is introduced in by (Lewis & Parke, 1987) that mouth shape changes according to extracted phonemes from a recorded speech. Their system, which could support control parameters for speech signal and visual parameters, is using software developed by (Pearce, et al., 1986). The control parameters allocated to mouth width, mouth corner x (horizontal), y (vertical), z offset, jaw rotation, mouth z (forward-back) offset relative to face, width of lips at mouth corner, teeth z and x offsets mouth corner x (horizontal), y (vertical), z offsets (with respect to rest of mouth), tapered lower lip and tapered upper lip raise relative to lower lip (Cohen & Massaro, 1993).

In this section, some visual speech animation methods are studied and categorized as image-based physically-based, parametric models, performance driven and statistical moles approaches. A seven minutes animation movie, in which speech became synchronized to facial animations, created and Lewis and Parke 1987 introduced system that is more sophisticated. Mouth shape changes according to extracted phonemes from a recorded speech (Pearce, et al., 1986) and Hill, et al in 1988. In 1990's, performance driven method, which human facial performances transfer to a synthetic face model, introduced. A physically base 3D facial animation is introduced in (Terzopoulos& Waters, 1990). This automated model consists of several facial layers like tissue, bone and muscles with control over facial muscles.

It can also estimates the facial muscle contractions from video sequences. More specifically skin tissue was simulated with three layers (epidermis, dermal and subcutaneous) that has elastic spring characteristics. In 1997, a method base on facial optical flow for animating facial components was introduced (Essa& Pentland, 1997). This method contributes to facial muscle deformation by creating facial expressions.

a. IMAGE-BASED

In this method, the visual speech was synthesized by selecting a number of key images (frames) usually less than the entire images covering the lip movement. Image-based modelling has been chosen to generate 3D face model. The proposed approach in (Williams, 1990) was to synthesis facial expressions from texture coordinates by tracking 2D feature points and the geometric and texture information are the subject of facial expression extracting (Guenter, et al., 1998). The approach in (Terzopoulos & Waters, 1990) was to capture facial expressions online and blend them to predefined 3D facial model to animate character.

b. MUSCLE-BASED (PHYSICS-BASED)

In this method, physical characteristics of skin and muscles are the subject of controlling and creating facial animations. Skin in this method has biomechanical properties and can be attached by muscle contraction. The virtual muscles simulate the relation between those factors using mass-spring and finite element algorithms. Using the biomechanical properties of skin and attaching the idea of virtual muscle to it that was able to counteract and affect those properties animation was created. The suggested methods to model the facial muscle in (Platt & Badler, 1981), (Lee, et al., 1993), (Choe, et al., 2001) and (Koch, et al., 1998) was to use mass-spring or finite element networks; in (Lee, et al., 1993) modelling mouth by a two-dimensional mass-spring in the territory of speech animation. The suggested method by (Platt & Badler, 1981) a structured model with skin, muscle, and bone nodes connected together via spring models.

c. PARAMETRIC METHOD

Parameterization method specifies the facial expressions by parametric values (Parke & Waters, 1996) providing control for desirable expressions. In this method mouth region divides to different parts to represent jaw rotation, mouth width, upper lip position and controlling mouth corners. This approach, by avoiding the complexity of specifying key frames for facial expression tries to approximate deformation of skin by defining a set of control parameters to control regions or features of face. The parametric model, which can be rendered by skin texture, describes the face geometry

with a set of 3D meshes. Between the skull points and skin feature points, which territory of effect has been defined, a set of muscles has been created. More specifically there are two type of control points are defined as expression parameters that controls mouth, eyebrows, etc. and conformation parameters that controls the topology of the face.

The MPEG-4 standard (ISO/IEC, 2001) was also using for the visual speech animation. A 3D prototype of face was allocated to facial animation by controlling two types of parameters as Facial definition parameters (FDPs) and Facial Animation Parameters (FAP). The FDPs can be used for generating various facial models while the Frame-based and DCT can be used for coding FAP in order to produce articulating lips. The number of the feature points located on the model was 84 and illustrated in Figure 8-2 which the lip was modelled by 18 feature points. The MPEG-4 also provides 14 visemes adopted from (Kshirsagar, et al., 1999) as it represented in (ISO/IEC, 2001).

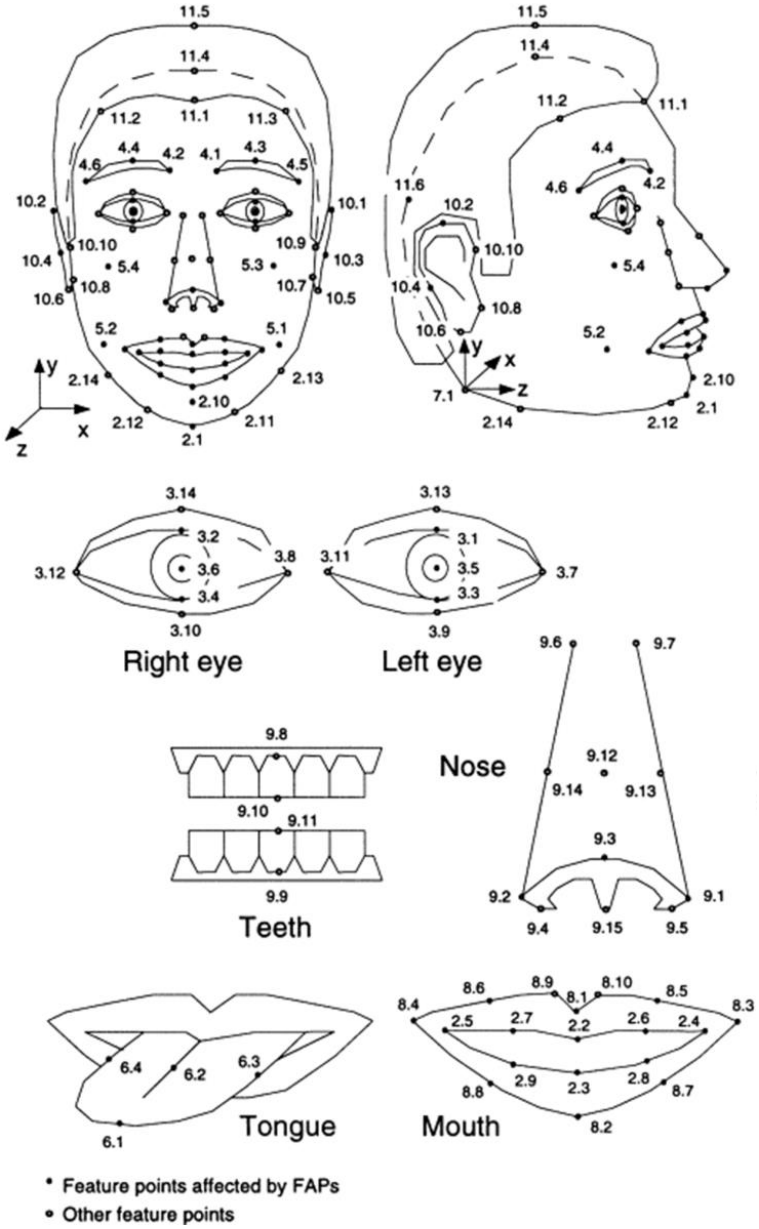


Figure 8-2: The facial feature points in the MPEG-4 standard (Lavagetto, et al., 2000)

d. PERFORMANCE-BASED

Capturing the dynamic of facial gesture was regarded by performance-based approach. The actual human face behaviour synthesizes a face by tracking markers located on face or tracking the facial features (lips, eyes, nose, etc.). This was an approach to control avatar-based models. In the proposed method by (Williams, 1990), (Lewis & Parke, 1987), for producing virtual model in motion, real faces are tracked. A synchronization method for animation and speech with linear prediction introduced in (Pearce, et al., 1986), used to generate a set of twelve reference

phonemes by analysing the speech signal. The parameters for lip shape and jaw rotation are associated to each phoneme. Therefore, key frames of parametric face model can be rendered.

B LOG LIKELIHOOD IN PARAMETRIC ESTIMATION

Table 8-5: The similarities between the families of distributions and the visual speech sample sets before and after processing in term of log likelihood

	Log likelihood					
	Sample sets befor modification			Sample sets after modification		
Distribution	$u^{W_m}[n]$	$l^{W_m}[n]$	$c^{W_m}[n]$	$u^{W_m}[n]$	$l^{W_m}[n]$	$c^{W_m}[n]$
Birnbaum-Saunders	-2463.03	-2736.91	-2554.89	-2399.43	-2613.29	-2240.63
Exponential	-3230.28	-3529.95	-3967.14	-3256.74	-3527.61	-3964.85
Gamma	-2598.07	-2730.51	-2556.65	-2402.58	-2614.66	-2241.77
Generalized extreme value	-2461.27	-2723.86	-2549.53	-2399.08	-2612.85	-2247.37
Inverse Gaussian	-1873.15	-2147.12	-1964.93	-1809.45	-2023.33	-1650.67
Log-Logistic	-2463.13	-2749.37	-2567.48	-2409.33	-2630.59	-2219.54
Logistic	-2465.71	-2742.89	-2574.49	-2420.45	-2634.51	-2223.21
Lognormal	-2462.71	-2737.27	-2554.93	-2399.57	-2613.7	-2240.54
Nakagami	-2464.83	-2726.69	-2558.85	-2408.58	-2617.6	-2243.38
Negative Binomial	-2462.05	-2729.27	-	-2405.66	-2615.99	-
Normal	-2473.01	-2726.85	-2561.62	-2418.04	-2623.39	-2245.2
Poisson	-2608.74	-3064.95	-2559.22	-2480.31	-2766.7	-2366.59
Rayleigh	-2824.1	-3119.19	-3527.33	-2841.75	-3107.27	-3521.81
Rician	-2472.55	-2726.75	-2561.6	-2417.67	-2623.2	-2245.2
t location-scale	-2465.08	-	-2561.62	-2416.49	-	-2219.76
Weibull	-2506.14	-2733.23	-2610.9	-2455.88	-2653.7	-2336.38

C VISEME GROUPS

Table 8-6: The phoneme-viseme mapping tables extended from (Ramage, 2011)

	Viseme groups																
(Heider & Heider, 1940)	/p, b, m/	/f, v/	/r/	/sh, ch, jh/	/n, t, d/	/th/	/k, g/	/l/									
(Woodward & Barber, 1960)	/p, b, m/	/f, v/	/w, r, hw/	/t, d, n, l, th, dh/	/s, z, ch, jh, sh, zh/	/j, k, g, h/											
(Fisher, 1968)	/p, b, m, d/	/f, v/	/w, hw, r/														
(Fisher, 1968)	/p, b/	/f, v/	/sh, zh, jh, ch/	/t, d, n, th, dh, s, z, r, l/	/z, r, l/												
(Binnie, et al., 1974)	/p, b, m/	/f, v/	/sh, zh/	/t, d, n, s, z, k, g/	/th/												
(Binnie, et al., 1976)	/p, b, m/	/f, v/	/r/	/w/	/sh, zh/	/t, d, s, z/	/th, dh/	/k, g/									
(Walden, et al., 1977)	/p, b, m/	/f,v/	/w/	/r/	/sh, zh/	/s, z/	/th, dh/	/t, d, n, k, g, j/	/l/								
(Walden, et al., 1981)	/p, b, m/	/f, v/	/w, r/	/sh, zh, ch, jh/	/th, dh/												
(Benguerel & Pichora-Fuller, 1982)	/p/	/f/	/w/	/ch, sh/	/th/												
(Garofolo, et al., 1993)	/b, d, g, p, t, k, dx, q/	/jh, ch/	/s, sh, z, zh, f, th, v, dh/	/m, n, ng, em, en, eng, nx/	/l, r, w, v, hh, hv/	/el/	/iy, ih, eh, ey, ae, aa, aw, ay, ah, ao, oy, ow, uh, uw, ux, er, ix, axr, ax-h/										
(Goldschen, et al., 1994)	/p, b/ /bcl, m, pcl/	/f, v/	/w/	/r/	/ch/ /jh/ /zh/	/sh, zh, jh, (ch)/	/t, d, n, th, dh, s/	z, r, l)/	/d, dcl, g, gcl, k, kcl, l, n, t, tcl/	/th/	/dh, epi/	/en/	/ng/	/dx, nx, q/	/hh/	/hv/	/y/
(Williams, 1990)	/p, b, m/	/f, v/	/r/	/l/	/sh, zh/	/th, dh/	/w/	/n, y, d, g, k, s, z, t/									

(Chen & Rao, 1998)	/p, b, m/	/f, v/	/w/	/r/	/sh, zh/	/d, g, k, t, y/	/th, dh/	/l/	/th/					
(Kshirsagar, et al., 1999) MPEG-4	/p, b, m/	/f, v/	/T, D/	/t, d/	/k, g/	/tS, dZ, S/	/s, z/	/n, l/	/r/	/A:/	/e/	/I/	/Q/	/U/
(Faruquie, et al., 2001)	/a, h/	/e, i/	/o, u/	/l/	/r/	/p, b, m/	/g, k, d, n, t, y/	/f, v, w/	/h, j, s, z/	/sh, ch/	/th/			
(Jiang, et al., 2001)	/m, p, b/	/f, v/	/r/	/w/	/, θ, , ð, /	/dʒ, ʒ, ʃ, tʃ/	/t, d, s, z/	/l, n/	/k, g, yh/					
(Lucey, et al., 2004)	/p, b, m, em/	/f, v/	/w, wh, r/	/ch, jh, sh, zh/	/k, g, n, l, nx, hh, y, el, en/	/iy, ih/	/aa/	/ah, ax, ay/	/eh, ey, ae, aw/	/uh, uw/	/ao, oy, ix, ow/	/er/		
(Saenko, et al., 2004)	/ax, ih, iy, dx/	/ah, aa/	/ae, eh, ay, ey, hh/	/aw, uh, u, wow, ao, w, oy/	/el, l/	/er, axr, r/	/Y/	/b, p/	/bcl, pcl, m, em/	/s, z, epi, tcl, dcl, n, en/	/ch, jh, sh, zh/	/t, d, th, dh, g, k/	/f, v/	/gcl, kcl, ng/
(Hazen, et al., 2004)	/b, p/ /bcl, pcl, m, em/	/f, v/	/w, aw, uh, uw, ow, ao, oy/	/r, er, axr/	/ch, jh, sh, zh/	/t, d, th, dh, g, k/	/gcl, kcl, ng/	/el, l/	/ah, aa/	/ax, ih, iy, dx/	/ae, eh, ay, ey, hh/	/y/	/th, dh/	
(Potamianos, et al., 2004)	/p, b, m/	/f, v/	/l, el, r, y/	/s, z/	/sh, zh, ch, jh/	/t, d, n, en/	/th, dh/	/ng, k, g, w/	/ih, iy, ax/	/ae, eh, ey, ay/	/uw, uh, ow/	/ao, ah, aa, er, oy, aw, hh/		
(Birkholz, et al., 2011)	/m, n, ŋ/	/a, ɛ, e, o, ø, æ, O/												

D THE ALGORITHMS

In this section, the algorithms for the Barycentric Lagrange Interpolation (BLI), barcode generation are shown in Figure 8-3 and Figure 8-4, respectively.

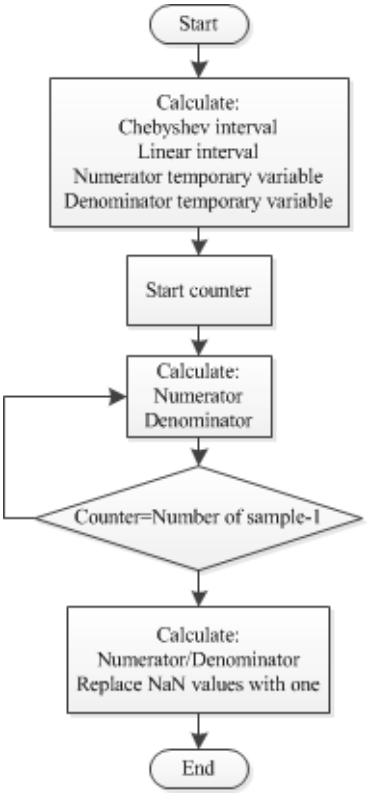


Figure 8-3: The algorithm for Barycentric Lagrange Interpolation (BLI)

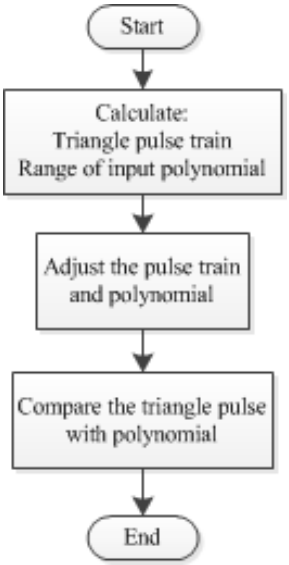


Figure 8-4: The algorithm for barcode generation

E THE MATHEMATICAL EXPRESSIONS OF VISUAL SPEECH SIGNALS

a. UPPER VISUAL SPEECH SIGNALS

$VS_{BLI}^{u\ W_1}(f_x)$=1./(3.6980.*(1./(fi-1)-1./(fi-2)+1./(fi-3)-1./(fi-4)+1./(fi-5)-1./(fi-6)+1./(fi-7)-1./(fi-8)+1./(fi-9)-1./(fi-10)+1./(fi-11)-1./(fi-12)+1./(fi-13)-1./(fi-14)+1./(fi-15)-1./(fi-16)+1./(fi-17)-1./(fi-18)+1./(fi-19)-1./(fi-20)+1./(fi-21)-1./(fi-22)+1./(fi-23)-1./(fi-24)+1./(fi-25)-1./(fi-26)+1./(fi-27)-1./(fi-28)+1./(fi-29)-1./(fi-30)+1./(fi-31)-1./(fi-32)+1./(fi-33)-1./(fi-34)+1./(fi-35)-1./(fi-36)+1./(fi-37)-1./(fi-38)+1./(fi-39)-1./(fi-40)+1./(fi-41)-1./(fi-42)+1./(fi-43)-1./(fi-44)+1./(fi-45)-1./(fi-46)+1./(fi-47)-1./(fi-48)+1./(fi-49)-1./(fi-50)+1./(fi-51)-1./(fi-52)+0.5./(fi-53)-0.5./fi)).*(1.9230e+002./(fi-1)-1.9599e+002./(fi-2)+1.9969e+002./(fi-3)-2.2928e+002./(fi-4)+2.4407e+002./(fi-5)-2.4407e+002./(fi-6)+2.4407e+002./(fi-7)-2.5516e+002./(fi-8)+2.4777e+002./(fi-9)-2.5147e+002./(fi-10)+2.4777e+002./(fi-11)-2.2558e+002./(fi-12)+2.2558e+002./(fi-13)-2.2188e+002./(fi-14)+2.1079e+002./(fi-15)-1.9969e+002./(fi-16)+1.9969e+002./(fi-17)-1.9599e+002./(fi-18)+1.9230e+002./(fi-19)-2.1079e+002./(fi-20)+2.0709e+002./(fi-21)-1.8120e+002./(fi-22)+1.7750e+002./(fi-23)-1.6641e+002./(fi-24)+1.6641e+002./(fi-25)-1.5532e+002./(fi-26)+1.4052e+002./(fi-27)-1.5162e+002./(fi-28)+1.5162e+002./(fi-29)-1.6271e+002./(fi-30)+1.7011e+002./(fi-31)-1.7011e+002./(fi-32)+1.8860e+002./(fi-33)-2.0709e+002./(fi-34)+2.1079e+002./(fi-35)-2.0709e+002./(fi-36)+2.0339e+002./(fi-37)-2.1079e+002./(fi-38)+2.0709e+002./(fi-39)-2.1079e+002./(fi-40)+2.0709e+002./(fi-41)-2.1079e+002./(fi-42)+2.2188e+002./(fi-43)-2.1448e+002./(fi-44)+2.0339e+002./(fi-45)-2.1079e+002./(fi-46)+2.1079e+002./(fi-47)-2.0709e+002./(fi-48)+1.9969e+002./(fi-49)-1.9599e+002./(fi-50)+1.8490e+002./(fi-51)-1.8860e+002./(fi-52)+9.0601e+001./(fi-53)-9.0601e+001./fi) (8-1)

$VS_{BLI}^{u\ W_2}(f_x)$=1./(8.5283.*(1./(fi-1)-1./(fi-2)+1./(fi-3)-1./(fi-4)+1./(fi-5)-1./(fi-6)+1./(fi-7)-1./(fi-8)+1./(fi-9)-1./(fi-10)+1./(fi-11)-1./(fi-12)+1./(fi-13)-1./(fi-14)+1./(fi-15)-1./(fi-16)+1./(fi-17)-1./(fi-18)+1./(fi-19)-1./(fi-20)+1./(fi-21)-1./(fi-22)+1./(fi-23)-1./(fi-24)+1./(fi-25)-1./(fi-26)+1./(fi-27)-1./(fi-28)+1./(fi-29)-1./(fi-30)+1./(fi-31)-1./(fi-32)+1./(fi-33)-1./(fi-34)+1./(fi-35)-1./(fi-36)+1./(fi-37)-1./(fi-38)+1./(fi-39)-1./(fi-40)+1./(fi-41)-1./(fi-42)+1./(fi-43)-1./(fi-44)+1./(fi-45)-1./(fi-46)+1./(fi-47)-1./(fi-48)+0.5./(fi-49)-0.5./fi)).*(4.1789e+002./(fi-1)-4.0936e+002./(fi-2)+4.4347e+002./(fi-3)-4.9464e+002./(fi-4)+5.1170e+002./(fi-5)-5.2875e+002./(fi-6)+5.2875e+002./(fi-7)-5.2875e+002./(fi-8)+5.2023e+002./(fi-9)-5.2023e+002./(fi-10)+5.0317e+002./(fi-11)-4.8611e+002./(fi-12)+4.8611e+002./(fi-13)-4.6906e+002./(fi-14)+4.5200e+002./(fi-15)-4.5200e+002./(fi-16)+4.7758e+002./(fi-17)-4.6906e+002./(fi-18)+4.5200e+002./(fi-19)-4.3494e+002./(fi-20)+4.2641e+002./(fi-21)-4.1789e+002./(fi-22)+3.6672e+002./(fi-23)-3.6672e+002./(fi-24)+3.6672e+002./(fi-25)-3.8377e+002./(fi-26)+3.9230e+002./(fi-27)-3.8377e+002./(fi-28)+4.0936e+002./(fi-29)-4.4347e+002./(fi-30)+4.7758e+002./(fi-31)-4.6906e+002./(fi-32)+4.7758e+002./(fi-33)-5.0317e+002./(fi-34)+5.2875e+002./(fi-35)-5.3728e+002./(fi-36)+5.3728e+002./(fi-37)-5.2875e+002./(fi-38)+5.1170e+002./(fi-39)-5.1170e+002./(fi-40)+5.2023e+002./(fi-41)-5.2023e+002./(fi-42)+5.0317e+002./(fi-43)-5.1170e+002./(fi-44)+4.9464e+002./(fi-45)-4.8611e+002./(fi-46)+4.6053e+002./(fi-47)-4.5200e+002./(fi-48)+2.0894e+002./(fi-49)-2.0894e+002./fi) (8-2)

$VS_{BLI}^{u\ W_3}(f_x)$=1./(3.6292.*(1./(fi-1)-1./(fi-2)+1./(fi-3)-1./(fi-4)+1./(fi-5)-1./(fi-6)+1./(fi-7)-1./(fi-8)+1./(fi-9)-1./(fi-10)+1./(fi-11)-1./(fi-12)+1./(fi-13)-1./(fi-14)+1./(fi-15)-1./(fi-16)+1./(fi-17)-1./(fi-18)+1./(fi-19)-1./(fi-20)+1./(fi-21)-1./(fi-22)+1./(fi-23)-1./(fi-24)+1./(fi-25)-1./(fi-26)+1./(fi-27)-1./(fi-28)+1./(fi-29)-1./(fi-30)+1./(fi-31)-1./(fi-32)+1./(fi-33)-1./(fi-34)+1./(fi-35)-1./(fi-36)+1./(fi-37)-1./(fi-38)+1./(fi-39)-1./(fi-40)+1./(fi-41)-1./(fi-42)+1./(fi-43)-1./(fi-44)+1./(fi-45)-1./(fi-46)+0.5./(fi-47)-0.5./fi)).*(1.7420e+002./(fi-1)-1.8509e+002./(fi-2)+1.9598e+002./(fi-3)-2.2864e+002./(fi-4)+2.3227e+002./(fi-5)-2.4679e+002./(fi-6)+2.5404e+002./(fi-7)-2.5404e+002./(fi-8)+2.5404e+002./(fi-9)-2.3953e+002./(fi-10)+2.2864e+002./(fi-11)-2.1775e+002./(fi-12)+2.1049e+002./(fi-13)-1.9961e+002./(fi-14)+1.9961e+002./(fi-15)-2.1049e+002./(fi-16)+1.9235e+002./(fi-17)-2.0323e+002./(fi-18)+1.6331e+002./(fi-19)-1.4880e+002./(fi-20)+1.3428e+002./(fi-21)-1.3428e+002./(fi-22)+1.3428e+002./(fi-23)-1.4880e+002./(fi-24)+1.5243e+002./(fi-25)-1.7057e+002./(fi-26)+1.7057e+002./(fi-27)-2.3953e+002./(fi-28)+2.5404e+002./(fi-29)-2.5404e+002./(fi-30)+2.4679e+002./(fi-31)-2.5041e+002./(fi-32)+2.4679e+002./(fi-33)-2.3590e+002./(fi-34)+2.3590e+002./(fi-35)-2.2864e+002./(fi-36)+2.2138e+002./(fi-37)-2.1412e+002./(fi-38)+2.1775e+002./(fi-39)-2.1412e+002./(fi-40)+2.0323e+002./(fi-41)-1.9961e+002./(fi-42)+1.9598e+002./(fi-43)-1.9961e+002./(fi-44)+1.9961e+002./(fi-45)-1.8509e+002./(fi-46)+8.8915e+001./(fi-47)-8.8915e+001./fi) (8-3)

$VS_{BLI}^{u\ W_3}(f_x)$=1./(3.6292.*(1./(fi-1)-1./(fi-2)+1./(fi-3)-1./(fi-4)+1./(fi-5)-1./(fi-6)+1./(fi-7)-1./(fi-8)+1./(fi-9)-1./(fi-10)+1./(fi-11)-1./(fi-12)+1./(fi-13)-1./(fi-14)+1./(fi-15)-1./(fi-16)+1./(fi-17)-1./(fi-18)+1./(fi-19)-1./(fi-20)+1./(fi-21)-1./(fi-22)+1./(fi-23)-1./(fi-24)+1./(fi-25)-1./(fi-26)+1./(fi-27)-1./(fi-28)+1./(fi-29)-1./(fi-30)+1./(fi-31)-1./(fi-32)+1./(fi-33)-1./(fi-34)+1./(fi-35)-1./(fi-36)+1./(fi-37)-1./(fi-38)+1./(fi-39)-1./(fi-40)+1./(fi-41)-1./(fi-42)+1./(fi-43)-1./(fi-44)+1./(fi-45)-1./(fi-46)+0.5./(fi-47)-0.5./fi)).*(1.7783e+002./(fi-1)-1.8509e+002./(fi-2)+1.9961e+002./(fi-3)-2.2864e+002./(fi-4)+2.5041e+002./(fi-5)-2.8671e+002./(fi-6)+2.8308e+002./(fi-7)-2.8671e+002./(fi-8)+2.9034e+002./(fi-9)-2.8671e+002./(fi-10)+2.8308e+002./(fi-11)-2.6493e+002./(fi-12)+2.5767e+002./(fi-13)-2.4316e+002./(fi-14)+2.4316e+002./(fi-15)-2.5041e+002./(fi-16)+2.4316e+002./(fi-17)-2.4316e+002./(fi-18)+2.4316e+002./(fi-19)-2.5767e+002./(fi-20)+2.5767e+002./(fi-21)-2.2501e+002./(fi-22)+1.8872e+002./(fi-23)-1.5606e+002./(fi-24)+1.4517e+002./(fi-25)-1.5606e+002./(fi-26)+1.5968e+002./(fi-27)-1.8146e+002./(fi-28)+2.0686e+002./(fi-29)-2.3590e+002./(fi-30)+2.3590e+002./(fi-31)-2.3953e+002./(fi-32)+2.3227e+002./(fi-33)-2.2501e+002./(fi-34)+2.1775e+002./(fi-35)-2.1775e+002./(fi-36)+2.0686e+002./(fi-37)-2.0686e+002./(fi-38)+2.0686e+002./(fi-39)-2.1049e+002./(fi-40)+2.0686e+002./(fi-41)-2.0323e+002./(fi-42)+2.0323e+002./(fi-43)-1.9235e+002./(fi-44)+1.9235e+002./(fi-45)-1.8872e+002./(fi-46)+8.8915e+001./(fi-47)-8.8915e+001./fi); (8-4)

$VS_{BLI}^{u\ W_5}(f_x)$=(9.1568e+001./(fi-1)-9.7291e+001./(fi-2)+9.3475e+001./(fi-3)-9.3475e+001./(fi-4)+9.5383e+001./(fi-5)-1.1446e+002./(fi-6)+1.2400e+002./(fi-7)-1.3735e+002./(fi-8)+1.4117e+002./(fi-9)-1.4307e+002./(fi-10)+1.4689e+002./(fi-11)-1.4498e+002./(fi-12)+1.4307e+002./(fi-13)-1.3735e+002./(fi-14)+1.3163e+002./(fi-15)-1.2400e+002./(fi-16)+1.2209e+002./(fi-17)-1.2972e+002./(fi-18)+1.1827e+002./(fi-19)-1.0492e+002./(fi-20)+9.5383e+001./(fi-21)-9.3475e+001./(fi-22)+9.5383e+001./(fi-23)-9.1568e+001./(fi-24)+9.3475e+001./(fi-25)-9.3475e+001./(fi-26)+9.3475e+001./(fi-27)-9.3475e+001./(fi-28)+1.0492e+002./(fi-29)-1.2209e+002./(fi-30)+1.2400e+002./(fi-31)-1.1827e+002./(fi-32)+1.1637e+002./(fi-33)-1.1827e+002./(fi-34)+1.2972e+002./(fi-35)-1.2972e+002./(fi-36)+1.2018e+002./(fi-37)-1.2209e+002./(fi-38)+1.2400e+002./(fi-39)-1.2972e+002./(fi-40)+1.2400e+002./(fi-41)-1.2400e+002./(fi-42)+1.2209e+002./(fi-43)-1.2400e+002./(fi-44)+1.2400e+002./(fi-45)-1.2781e+002./(fi-46)+1.2972e+002./(fi-47)-1.2400e+002./(fi-48)+1.2209e+002./(fi-49)-1.2209e+002./(fi-50)+1.2400e+002./(fi-51)-1.1637e+002./(fi-52)+1.1827e+002./(fi-53)-1.1255e+002./(fi-54)+1.0683e+002./(fi-55)-1.0874e+002./(fi-56)+1.0111e+002./(fi-57)-9.7291e+001./(fi-58)+4.6738e+001./(fi-59)-4.6738e+001./fi)./(1.9077.*(1./(fi-1)-1./(fi-2)+1./(fi-3)-1./(fi-4)+1./(fi-5)-1./(fi-6)+1./(fi-7)-1./(fi-8)+1./(fi-9)-1./(fi-10)+1./(fi-11)-1./(fi-12)+1./(fi-13)-1./(fi-14)+1./(fi-15)-1./(fi-16)+1./(fi-17)-1./(fi-18)+1./(fi-19)-1./(fi-20)+1./(fi-21)-1./(fi-22)+1./(fi-23)-1./(fi-24)+1./(fi-25)-1./(fi-26)+1./(fi-27)-1./(fi-28)+1./(fi-29)-1./(fi-30)+1./(fi-31)-1./(fi-32)+1./(fi-33)-1./(fi-34)+1./(fi-35)-1./(fi-36)+1./(fi-37)-1./(fi-38)+1./(fi-39)-1./(fi-40)+1./(fi-41)-1./(fi-42)+1./(fi-43)-1./(fi-44)+1./(fi-45)-1./(fi-46)+1./(fi-47)-1./(fi-48)+1./(fi-49)-1./(fi-50)+1./(fi-51)-1./(fi-52)+1./(fi-53)-1./(fi-54)+1./(fi-55)-1./(fi-56)+1./(fi-57)-1./(fi-58)+0.5./(fi-59)-0.5./fi)); (8-5)

$VS_{BLI}^{u\ W_6}(f_x)$=1./(5.0716*(1./(fi-1)-1./(fi-2)+1./(fi-3)-1./(fi-4)+1./(fi-5)-1./(fi-6)+1./(fi-7)-1./(fi-8)+1./(fi-9)-1./(fi-10)+1./(fi-11)-1./(fi-12)+1./(fi-13)-1./(fi-14)+1./(fi-15)-1./(fi-16)+1./(fi-17)-1./(fi-18)+1./(fi-19)-1./(fi-20)+1./(fi-21)-1./(fi-22)+1./(fi-23)-1./(fi-24)+1./(fi-25)-1./(fi-26)+1./(fi-27)-1./(fi-28)+1./(fi-29)-1./(fi-30)+1./(fi-31)-1./(fi-32)+1./(fi-33)-1./(fi-34)+1./(fi-35)-1./(fi-36)+1./(fi-37)-1./(fi-38)+1./(fi-39)-1./(fi-40)+1./(fi-41)-1./(fi-42)+0.5./(fi-43)-0.5./fi)).*(2.5865e+002./(fi-1)-2.4851e+002./(fi-2)+2.8401e+002./(fi-3)-2.8401e+002./(fi-4)+3.3473e+002./(fi-5)-3.8037e+002./(fi-6)+4.7166e+002./(fi-7)-5.0716e+002./(fi-8)+4.9195e+002./(fi-9)-4.9195e+002./(fi-10)+4.5137e+002./(fi-11)-4.0573e+002./(fi-12)+3.9559e+002./(fi-13)-3.7023e+002./(fi-14)+3.5501e+002./(fi-15)-3.7023e+002./(fi-16)+3.5501e+002./(fi-17)-3.2965e+002./(fi-18)+3.0937e+002./(fi-19)-2.6880e+002./(fi-20)+2.1301e+002./(fi-21)-2.2822e+002./(fi-22)+1.9272e+002./(fi-23)-1.9272e+002./(fi-24)+2.0794e+002./(fi-25)-2.8908e+002./(fi-26)+3.5501e+002./(fi-27)-3.7023e+002./(fi-28)+3.6516e+002./(fi-29)-3.6008e+002./(fi-30)+3.8037e+002./(fi-31)-4.0573e+002./(fi-32)+4.0573e+002./(fi-33)-4.0573e+002./(fi-34)+4.4123e+002./(fi-35)-4.2602e+002./(fi-36)+3.9051e+002./(fi-37)-3.4487e+002./(fi-38)+2.8401e+002./(fi-39)-2.9415e+002./(fi-40)+2.4851e+002./(fi-41)-2.5358e+002./(fi-42)+1.2425e+002./(fi-43)-1.2425e+002./fi); (8-6)

$VS_{BLI}^{u}{}^{W_7}(\square_x)$=(4.4607e+002./(fi-2)-4.4607e+002./(fi-1)-4.3749e+002./(fi-3)+4.2892e+002./(fi-4)-4.8896e+002./(fi-5)+5.4901e+002./(fi-6)-5.7475e+002./(fi-7)+6.0048e+002./(fi-8)-6.1764e+002./(fi-9)+6.2622e+002./(fi-10)-6.2622e+002./(fi-11)+6.0048e+002./(fi-12)-5.8333e+002./(fi-13)+5.8333e+002./(fi-14)-5.4043e+002./(fi-15)+5.4043e+002./(fi-16)-5.4043e+002./(fi-17)+5.2328e+002./(fi-18)-5.4901e+002./(fi-19)+5.4901e+002./(fi-20)-5.4043e+002./(fi-21)+5.3186e+002./(fi-22)-5.0612e+002./(fi-23)+5.3186e+002./(fi-24)-4.2892e+002./(fi-25)+3.9460e+002./(fi-26)-3.6887e+002./(fi-27)+3.8602e+002./(fi-28)-3.6887e+002./(fi-29)+4.2034e+002./(fi-30)-5.1470e+002./(fi-31)+5.3186e+002./(fi-32)-5.8333e+002./(fi-33)+5.8333e+002./(fi-34)-6.1764e+002./(fi-35)+6.3480e+002./(fi-36)-6.4337e+002./(fi-37)+6.3480e+002./(fi-38)-5.7475e+002./(fi-39)+4.7181e+002./(fi-40)-4.7181e+002./(fi-41)+4.5465e+002./(fi-42)-4.5465e+002./(fi-43)+2.1017e+002./(fi-44)+2.1017e+002./fi)./(8.5783*(1./(fi-2)-1./(fi-1)-1./(fi-3)+1./(fi-4)-1./(fi-5)+1./(fi-6)-1./(fi-7)+1./(fi-8)-1./(fi-9)+1./(fi-10)-1./(fi-11)+1./(fi-12)-1./(fi-13)+1./(fi-14)-1./(fi-15)+1./(fi-16)-1./(fi-17)+1./(fi-18)-1./(fi-19)+1./(fi-20)-1./(fi-21)+1./(fi-22)-1./(fi-23)+1./(fi-24)-1./(fi-25)+1./(fi-26)-1./(fi-27)+1./(fi-28)-1./(fi-29)+1./(fi-30)-1./(fi-31)+1./(fi-32)-1./(fi-33)+1./(fi-34)-1./(fi-35)+1./(fi-36)-1./(fi-37)+1./(fi-38)-1./(fi-39)+1./(fi-40)-1./(fi-41)+1./(fi-42)-1./(fi-43)+0.5./(fi-44)+0.5./fi)); (8-7)

$VS_{BLI}^{u}{}^{W_8}(f_x)$=1./(1.8469.*(1./(fi-1)-1./(fi-2)+1./(fi-3)-1./(fi-4)+1./(fi-5)-1./(fi-6)+1./(fi-7)-1./(fi-8)+1./(fi-9)-1./(fi-10)+1./(fi-11)-1./(fi-12)+1./(fi-13)-1./(fi-14)+1./(fi-15)-1./(fi-16)+1./(fi-17)-1./(fi-18)+1./(fi-19)-1./(fi-20)+1./(fi-21)-1./(fi-22)+1./(fi-23)-1./(fi-24)+1./(fi-25)-1./(fi-26)+1./(fi-27)-1./(fi-28)+1./(fi-29)-1./(fi-30)+1./(fi-31)-1./(fi-32)+1./(fi-33)-1./(fi-34)+1./(fi-35)-1./(fi-36)+1./(fi-37)-1./(fi-38)+1./(fi-39)-1./(fi-40)+1./(fi-41)-1./(fi-42)+1./(fi-43)-1./(fi-44)+1./(fi-45)-1./(fi-46)+1./(fi-47)-1./(fi-48)+1./(fi-49)-1./(fi-50)+0.5./(fi-51)-0.5./fi)).*(9.0500e+001./(fi-1)-9.4194e+001./(fi-2)+9.9734e+001./(fi-3)-1.1082e+002./(fi-4)+1.1820e+002./(fi-5)-1.2005e+002./(fi-6)+1.2559e+002./(fi-7)-1.3667e+002./(fi-8)+1.3667e+002./(fi-9)-1.3483e+002./(fi-10)+1.2559e+002./(fi-11)-1.2374e+002./(fi-12)+1.1082e+002./(fi-13)-1.0897e+002./(fi-14)+1.0712e+002./(fi-15)-1.0158e+002./(fi-16)+9.0500e+001./(fi-17)-8.6806e+001./(fi-18)+8.3112e+001./(fi-19)-8.3112e+001./(fi-20)+7.2030e+001./(fi-21)-7.0183e+001./(fi-22)+7.3877e+001./(fi-23)-8.4959e+001./(fi-24)+9.2347e+001./(fi-25)-9.0500e+001./(fi-26)+9.4194e+001./(fi-27)-8.8653e+001./(fi-28)+8.8653e+001./(fi-29)-9.0500e+001./(fi-30)+1.1266e+002./(fi-31)-1.2190e+002./(fi-32)+1.4406e+002./(fi-33)-1.4960e+002./(fi-34)+1.4221e+002./(fi-35)-1.3852e+002./(fi-36)+1.2559e+002./(fi-37)-1.2374e+002./(fi-38)+1.1820e+002./(fi-39)-1.1266e+002./(fi-40)+9.9734e+001./(fi-41)-9.6040e+001./(fi-42)+9.2347e+001./(fi-43)-9.0500e+001./(fi-44)+9.0500e+001./(fi-45)-8.4959e+001./(fi-46)+9.2347e+001./(fi-47)-8.3112e+001./(fi-48)+8.4959e+001./(fi-49)-8.3112e+001./(fi-50)+4.5250e+001./(fi-51)-4.5250e+001./fi); (8-8)

$VS_{BLI}^{u}{}^{W_9}(f_x)$=1./(1.8469*(1./(fi-1)-1./(fi-2)+1./(fi-3)-1./(fi-4)+1./(fi-5)-1./(fi-6)+1./(fi-7)-1./(fi-8)+1./(fi-9)-1./(fi-10)+1./(fi-11)-1./(fi-12)+1./(fi-13)-1./(fi-14)+1./(fi-15)-1./(fi-16)+1./(fi-17)-1./(fi-18)+1./(fi-19)-1./(fi-20)+1./(fi-21)-1./(fi-22)+1./(fi-23)-1./(fi-24)+1./(fi-25)-1./(fi-26)+1./(fi-27)-1./(fi-28)+1./(fi-29)-1./(fi-30)+1./(fi-31)-1./(fi-32)+1./(fi-33)-1./(fi-34)+1./(fi-35)-1./(fi-36)+1./(fi-37)-1./(fi-38)+1./(fi-39)-1./(fi-40)+1./(fi-41)-1./(fi-42)+1./(fi-43)-1./(fi-44)+1./(fi-45)-1./(fi-46)+1./(fi-47)-1./(fi-48)+1./(fi-49)-1./(fi-50)+0.5./(fi-51)-0.5./fi)).*(9.0500e+001./(fi-1)-9.2347e+001./(fi-2)+9.2347e+001./(fi-3)-1.0528e+002./(fi-4)+1.1451e+002./(fi-5)-1.2559e+002./(fi-6)+1.2744e+002./(fi-7)-1.2190e+002./(fi-8)+1.3483e+002./(fi-9)-1.2744e+002./(fi-10)+1.2744e+002./(fi-11)-1.2005e+002./(fi-12)+1.2005e+002./(fi-13)-1.1082e+002./(fi-14)+1.1082e+002./(fi-15)-1.1082e+002./(fi-16)+1.1082e+002./(fi-17)-9.0500e+001./(fi-18)+7.7571e+001./(fi-19)-7.5724e+001./(fi-20)+8.4959e+001./(fi-21)-1.1266e+002./(fi-22)+1.0712e+002./(fi-23)-1.1266e+002./(fi-24)+1.1451e+002./(fi-25)-1.1266e+002./(fi-26)+1.0343e+002./(fi-27)-1.0343e+002./(fi-28)+9.7887e+001./(fi-29)-9.7887e+001./(fi-30)+1.0528e+002./(fi-31)-1.0712e+002./(fi-32)+1.1266e+002./(fi-33)-1.0158e+002./(fi-34)+1.0158e+002./(fi-35)-8.8653e+001./(fi-36)+8.3112e+001./(fi-37)-8.6806e+001./(fi-38)+8.4959e+001./(fi-39)-8.4959e+001./(fi-40)+8.4959e+001./(fi-41)-8.8653e+001./(fi-42)+8.8653e+001./(fi-43)-8.6806e+001./(fi-44)+8.6806e+001./(fi-45)-9.0500e+001./(fi-46)+9.0500e+001./(fi-47)-9.2347e+001./(fi-48)+8.8653e+001./(fi-49)-9.2347e+001./(fi-50)+4.5250e+001./(fi-51)-4.5250e+001./fi); (8-9)

$VS_{BLI}^{u}{}^{W_{10}}(f_x)$=(1.7750e+002./(fi-1)-1.8490e+002./(fi-2)+1.8860e+002./(fi-3)-2.1448e+002./(fi-4)+2.2188e+002./(fi-5)-2.2188e+002./(fi-6)+2.4037e+002./(fi-7)-2.6256e+002./(fi-8)+2.8475e+002./(fi-9)-2.9584e+002./(fi-10)+2.8845e+002./(fi-11)-2.9584e+002./(fi-12)+2.8845e+002./(fi-13)-2.7365e+002./(fi-14)+2.5516e+002./(fi-15)-2.4407e+002./(fi-16)+2.4037e+002./(fi-17)-2.4037e+002./(fi-18)+2.4037e+002./(fi-19)-2.4037e+002./(fi-20)+2.4037e+002./(fi-21)-2.2188e+002./(fi-22)+2.0339e+002./(fi-23)-1.8120e+002./(fi-24)+1.7750e+002./(fi-25)-1.6641e+002./(fi-26)+1.8120e+002./(fi-27)-2.1448e+002./(fi-28)+2.2558e+002./(fi-29)-2.5516e+002./(fi-30)+2.8105e+002./(fi-31)-3.1063e+002./(fi-32)+3.2173e+002./(fi-33)-3.2912e+002./(fi-34)+3.1803e+002./(fi-35)-3.1803e+002./(fi-36)+3.0694e+002./(fi-37)-2.8475e+002./(fi-38)+2.6626e+002./(fi-39)-2.4777e+002./(fi-40)+2.2558e+002./(fi-41)-2.2558e+002./(fi-42)+2.1079e+002./(fi-43)-2.0709e+002./(fi-44)+2.1079e+002./(fi-45)-2.0339e+002./(fi-46)+2.0709e+002./(fi-47)-2.0709e+002./(fi-48)+2.0709e+002./(fi-49)-1.9599e+002./(fi-50)+1.8490e+002./(fi-51)-1.8860e+002./(fi-52)+9.0601e+001./(fi-53)-9.0601e+001./fi)./(3.6980*(1./(fi-1)-1./(fi-2)+1./(fi-3)-1./(fi-4)+1./(fi-5)-1./(fi-6)+1./(fi-7)-1./(fi-8)+1./(fi-9)-1./(fi-10)+1./(fi-11)-1./(fi-12)+1./(fi-13)-1./(fi-14)+1./(fi-15)-1./(fi-16)+1./(fi-17)-1./(fi-18)+1./(fi-19)-1./(fi-20)+1./(fi-21)-1./(fi-22)+1./(fi-23)-1./(fi-24)+1./(fi-25)-1./(fi-26)+1./(fi-27)-1./(fi-28)+1./(fi-29)-1./(fi-30)+1./(fi-31)-1./(fi-32)+1./(fi-33)-1./(fi-34)+1./(fi-35)-1./(fi-36)+1./(fi-37)-1./(fi-38)+1./(fi-39)-1./(fi-40)+1./(fi-41)-1./(fi-42)+1./(fi-43)-1./(fi-44)+1./(fi-45)-1./(fi-46)+1./(fi-47)-1./(fi-48)+1./(fi-49)-1./(fi-50)+1./(fi-51)-1./(fi-52)+0.5./(fi-53)-0.5./fi)); (8-10)

$VS_{BLI}^{u}{}^{W_{11}}(f_x)$=1./(8.5283*(1./(fi-1)-1./(fi-2)+1./(fi-3)-1./(fi-4)+1./(fi-5)-1./(fi-6)+1./(fi-7)-1./(fi-8)+1./(fi-9)-1./(fi-10)+1./(fi-11)-1./(fi-12)+1./(fi-13)-1./(fi-14)+1./(fi-15)-1./(fi-16)+1./(fi-17)-1./(fi-18)+1./(fi-19)-1./(fi-20)+1./(fi-21)-1./(fi-22)+1./(fi-23)-1./(fi-24)+1./(fi-25)-1./(fi-26)+1./(fi-27)-1./(fi-28)+1./(fi-29)-1./(fi-30)+1./(fi-31)-1./(fi-32)+1./(fi-33)-1./(fi-34)+1./(fi-35)-1./(fi-36)+1./(fi-37)-1./(fi-38)+1./(fi-39)-1./(fi-40)+1./(fi-41)-1./(fi-42)+1./(fi-43)-1./(fi-44)+1./(fi-45)-1./(fi-46)+1./(fi-47)-1./(fi-48)+0.5./(fi-49)-0.5./fi)).*(4.2641e+002./(fi-1)-4.0936e+002./(fi-2)+4.2641e+002./(fi-3)-4.2641e+002./(fi-4)+4.2641e+002./(fi-5)-4.8611e+002./(fi-6)+5.1170e+002./(fi-7)-5.7992e+002./(fi-8)+5.4581e+002./(fi-9)-5.7140e+002./(fi-10)+5.4581e+002./(fi-11)-5.7992e+002./(fi-12)+5.3728e+002./(fi-13)-5.5434e+002./(fi-14)+5.1170e+002./(fi-15)-4.9464e+002./(fi-16)+5.2023e+002./(fi-17)-5.5434e+002./(fi-18)+5.4581e+002./(fi-19)-4.4347e+002./(fi-20)+3.6672e+002./(fi-21)-3.6672e+002./(fi-22)+3.5819e+002./(fi-23)-3.7524e+002./(fi-24)+4.0936e+002./(fi-25)-3.9230e+002./(fi-26)+5.0317e+002./(fi-27)-5.0317e+002./(fi-28)+3.9230e+002./(fi-29)-5.5434e+002./(fi-30)+5.8845e+002./(fi-31)-5.2875e+002./(fi-32)+5.1170e+002./(fi-33)-5.2875e+002./(fi-34)+5.0317e+002./(fi-35)-4.8611e+002./(fi-36)+4.7758e+002./(fi-37)-4.8611e+002./(fi-38)+4.5200e+002./(fi-39)-4.6053e+002./(fi-40)+4.3494e+002./(fi-41)-4.4347e+002./(fi-42)+4.6906e+002./(fi-43)-4.5200e+002./(fi-44)+4.3494e+002./(fi-45)-4.2641e+002./(fi-46)+4.3494e+002./(fi-47)-4.0083e+002./(fi-48)+2.0894e+002./(fi-49)-2.0894e+002./fi); (8-11)

$VS_{BLI}^{u}{}^{W_{12}}(f_x)$=(4.8203e+002./(fi-2)-4.8203e+002./(fi-1)-5.0022e+002./(fi-3)+5.3660e+002./(fi-4)-5.5479e+002./(fi-5)+5.8208e+002./(fi-6)-6.1846e+002./(fi-7)+6.2755e+002./(fi-8)-6.4574e+002./(fi-9)+6.0936e+002./(fi-10)-6.0027e+002./(fi-11)+6.1846e+002./(fi-12)-6.5484e+002./(fi-13)+6.7303e+002./(fi-14)-6.8212e+002./(fi-15)+6.9122e+002./(fi-16)-7.0941e+002./(fi-17)+7.3669e+002./(fi-18)-7.2760e+002./(fi-19)+5.5479e+002./(fi-20)-5.6389e+002./(fi-21)+5.0932e+002./(fi-22)-4.7294e+002./(fi-23)+5.4570e+002./(fi-24)-6.0027e+002./(fi-25)+6.2755e+002./(fi-26)-6.5484e+002./(fi-27)+6.7303e+002./(fi-28)-6.5484e+002./(fi-29)+6.5484e+002./(fi-30)-6.2755e+002./(fi-31)+6.0936e+002./(fi-32)-5.9117e+002./(fi-33)+5.7298e+002./(fi-34)-5.4570e+002./(fi-35)+5.1841e+002./(fi-36)-5.0022e+002./(fi-37)+4.9113e+002./(fi-38)-4.7294e+002./(fi-39)+2.2283e+002./(fi-40)+2.2283e+002./fi)./(9.0949*(1./(fi-2)-1./(fi-1)-1./(fi-3)+1./(fi-4)-1./(fi-5)+1./(fi-6)-1./(fi-7)+1./(fi-8)-1./(fi-9)+1./(fi-10)-1./(fi-11)+1./(fi-12)-1./(fi-13)+1./(fi-14)-1./(fi-15)+1./(fi-16)-1./(fi-17)+1./(fi-18)-1./(fi-19)+1./(fi-20)-1./(fi-21)+1./(fi-22)-1./(fi-23)+1./(fi-24)-1./(fi-25)+1./(fi-26)-1./(fi-27)+1./(fi-28)-1./(fi-29)+1./(fi-30)-1./(fi-31)+1./(fi-32)-1./(fi-33)+1./(fi-34)-1./(fi-35)+1./(fi-36)-1./(fi-37)+1./(fi-38)-1./(fi-39)+0.5./(fi-40)+0.5./fi)); (8-12)

$VS_{BLI}^{u}{}^{W_{13}}(f_x)$=(2.5358e+002./(fi-1)-2.5865e+002./(fi-2)+2.7387e+002./(fi-3)-2.9415e+002./(fi-4)+3.0430e+002./(fi-5)-3.0430e+002./(fi-6)+3.1951e+002./(fi-7)-3.0937e+002./(fi-8)+3.2458e+002./(fi-9)-3.2458e+002./(fi-10)+3.0430e+002./(fi-11)-2.9415e+002./(fi-12)+2.9415e+002./(fi-13)-3.2458e+002./(fi-14)+3.0430e+002./(fi-15)-2.2315e+002./(fi-16)+1.9779e+002./(fi-17)-2.0794e+002./(fi-18)+2.1301e+002./(fi-19)-2.0286e+002./(fi-20)+2.8401e+002./(fi-21)-3.0430e+002./(fi-22)+2.9923e+002./(fi-23)-2.9923e+002./(fi-24)+2.9415e+002./(fi-25)-2.8401e+002./(fi-26)+2.8401e+002./(fi-27)-2.7894e+002./(fi-28)+2.7894e+002./(fi-29)-2.7387e+002./(fi-30)+2.7894e+002./(fi-31)-2.8908e+002./(fi-32)+2.7894e+002./(fi-33)-2.6880e+002./(fi-34)+2.6880e+002./(fi-35)-2.7387e+002./(fi-36)+2.6372e+002./(fi-37)-2.6372e+002./(fi-38)+2.4851e+002./(fi-39)-2.4851e+002./(fi-40)+2.4344e+002./(fi-41)-2.4851e+002./(fi-42)+1.2425e+002./(fi-43)-1.2425e+002./fi)./(5.0716*(1./(fi-1)-1./(fi-2)+1./(fi-3)-1./(fi-4)+1./(fi-5)-1./(fi-6)+1./(fi-7)-1./(fi-8)+1./(fi-9)-1./(fi-10)+1./(fi-11)-1./(fi-12)+1./(fi-13)-1./(fi-14)+1./(fi-15)-1./(fi-16)+1./(fi-17)-1./(fi-18)+1./(fi-19)-1./(fi-20)+1./(fi- (8-13)

21)-1./(fi-22)+1./(fi-23)-1./(fi-24)+1./(fi-25)-1./(fi-26)+1./(fi-27)-1./(fi-28)+1./(fi-29)-1./(fi-30)+1./(fi-31)-1./(fi-32)+1./(fi-33)-1./(fi-34)+1./(fi-35)-1./(fi-36)+1./(fi-37)-1./(fi-38)+1./(fi-39)-1./(fi-40)+1./(fi-41)-1./(fi-42)+0.5./(fi-43)-0.5./fi));

b. LOWER VISUAL SPEECH SIGNALS

$VS_{BLI}^{l\ W_1}(f_x)$=(2.5886e+002./(fi-1)-2.6256e+002./(fi-2)+2.8845e+002./(fi-3)-3.5501e+002./(fi-4)+3.7350e+002./(fi-5)-3.9939e+002./(fi-6)+4.2157e+002./(fi-7)-4.1788e+002./(fi-8)+4.3267e+002./(fi-9)-4.2157e+002./(fi-10)+4.1048e+002./(fi-11)-4.1048e+002./(fi-12)+3.9569e+002./(fi-13)-3.9569e+002./(fi-14)+3.8829e+002./(fi-15)-3.6610e+002./(fi-16)+3.6241e+002./(fi-17)-3.5501e+002./(fi-18)+3.6610e+002./(fi-19)-3.5871e+002./(fi-20)+3.5871e+002./(fi-21)-3.3652e+002./(fi-22)+3.3652e+002./(fi-23)-3.4761e+002./(fi-24)+3.5871e+002./(fi-25)-3.7350e+002./(fi-26)+3.5871e+002./(fi-27)-3.1063e+002./(fi-28)+2.9954e+002./(fi-29)-3.2173e+002./(fi-30)+3.4392e+002./(fi-31)-3.4392e+002./(fi-32)+3.7350e+002./(fi-33)-3.8459e+002./(fi-34)+3.6241e+002./(fi-35)-3.5131e+002./(fi-36)+3.4761e+002./(fi-37)-3.4022e+002./(fi-38)+3.4392e+002./(fi-39)-3.3282e+002./(fi-40)+3.4022e+002./(fi-41)-3.2912e+002./(fi-42)+3.1803e+002./(fi-43)-3.0694e+002./(fi-44)+3.0694e+002./(fi-45)-2.8845e+002./(fi-46)+2.8475e+002./(fi-47)-2.7735e+002./(fi-48)+2.7365e+002./(fi-49)-2.7735e+002./(fi-50)+2.7365e+002./(fi-51)-2.5147e+002./(fi-52)+1.2943e+002./(fi-53)-1.2943e+002./ffi)./(3.6980*(1./(fi-1)-1./(fi-2)+1./(fi-3)-1./(fi-4)+1./(fi-5)-1./(fi-6)+1./(fi-7)-1./(fi-8)+1./(fi-9)-1./(fi-10)+1./(fi-11)-1./(fi-12)+1./(fi-13)-1./(fi-14)+1./(fi-15)-1./(fi-16)+1./(fi-17)-1./(fi-18)+1./(fi-19)-1./(fi-20)+1./(fi-21)-1./(fi-22)+1./(fi-23)-1./(fi-24)+1./(fi-25)-1./(fi-26)+1./(fi-27)-1./(fi-28)+1./(fi-29)-1./(fi-30)+1./(fi-31)-1./(fi-32)+1./(fi-33)-1./(fi-34)+1./(fi-35)-1./(fi-36)+1./(fi-37)-1./(fi-38)+1./(fi-39)-1./(fi-40)+1./(fi-41)-1./(fi-42)+1./(fi-43)-1./(fi-44)+1./(fi-45)-1./(fi-46)+1./(fi-47)-1./(fi-48)+1./(fi-49)-1./(fi-50)+1./(fi-51)-1./(fi-52)+.5./(fi-53)-0.5./ffi)); (8-14)

$VS_{BLI}^{l\ W_2}(f_x)$=1./(8.5283*(1./(fi-1)-1./(fi-2)+1./(fi-3)-1./(fi-4)+1./(fi-5)-1./(fi-6)+1./(fi-7)-1./(fi-8)+1./(fi-9)-1./(fi-10)+1./(fi-11)-1./(fi-12)+1./(fi-13)-1./(fi-14)+1./(fi-15)-1./(fi-16)+1./(fi-17)-1./(fi-18)+1./(fi-19)-1./(fi-20)+1./(fi-21)-1./(fi-22)+1./(fi-23)-1./(fi-24)+1./(fi-25)-1./(fi-26)+1./(fi-27)-1./(fi-28)+1./(fi-29)-1./(fi-30)+1./(fi-31)-1./(fi-32)+1./(fi-33)-1./(fi-34)+1./(fi-35)-1./(fi-36)+1./(fi-37)-1./(fi-38)+1./(fi-39)-1./(fi-40)+1./(fi-41)-1./(fi-42)+1./(fi-43)-1./(fi-44)+1./(fi-45)-1./(fi-46)+1./(fi-47)-1./(fi-48)+0.5./(fi-49)-0.5./ffi)).*(6.1404e+002./(fi-1)-6.5668e+002./(fi-2)+6.9932e+002./(fi-3)-7.6755e+002./(fi-4)+8.2724e+002./(fi-5)-8.5283e+002./(fi-6)+8.6989e+002./(fi-7)-8.7841e+002./(fi-8)+8.6989e+002./(fi-9)-8.6136e+002./(fi-10)+8.6136e+002./(fi-11)-8.4430e+002./(fi-12)+8.4430e+002./(fi-13)-8.3577e+002./(fi-14)+8.4430e+002./(fi-15)-8.4430e+002./(fi-16)+8.2724e+002./(fi-17)-8.1019e+002./(fi-18)+7.6755e+002./(fi-19)-7.9313e+002./(fi-20)+7.5902e+002./(fi-21)-7.5902e+002./(fi-22)+7.2490e+002./(fi-23)-7.1638e+002./(fi-24)+7.7607e+002./(fi-25)-8.0166e+002./(fi-26)+7.9313e+002./(fi-27)-7.1638e+002./(fi-28)+7.5902e+002./(fi-29)-7.6755e+002./(fi-30)+7.7607e+002./(fi-31)-7.6755e+002./(fi-32)+7.8460e+002./(fi-33)-8.1019e+002./(fi-34)+7.7607e+002./(fi-35)-7.3343e+002./(fi-36)+7.2490e+002./(fi-37)-7.0785e+002./(fi-38)+7.1638e+002./(fi-39)-7.0785e+002./(fi-40)+6.9932e+002./(fi-41)-6.9079e+002./(fi-42)+6.8226e+002./(fi-43)-6.5668e+002./(fi-44)+6.4815e+002./(fi-45)-6.3962e+002./(fi-46)+6.2257e+002./(fi-47)-6.1404e+002./(fi-48)+2.9849e+002./(fi-49)-2.9849e+002./ffi); (8-15)

$VS_{BLI}^{l\ W_3}(f_x)$=(2.5041e+002./(fi-1)-2.5041e+002./(fi-2)+2.5767e+002./(fi-3)-3.2300e+002./(fi-4)+3.1937e+002./(fi-5)-3.3752e+002./(fi-6)+3.4840e+002./(fi-7)-3.5929e+002./(fi-8)+3.3752e+002./(fi-9)-3.3752e+002./(fi-10)+3.3026e+002./(fi-11)-3.3026e+002./(fi-12)+3.2663e+002./(fi-13)-3.3389e+002./(fi-14)+3.3026e+002./(fi-15)-3.3026e+002./(fi-16)+3.5203e+002./(fi-17)-3.5929e+002./(fi-18)+3.4114e+002./(fi-19)-3.1937e+002./(fi-20)+2.7582e+002./(fi-21)-2.3590e+002./(fi-22)+2.3590e+002./(fi-23)-2.2501e+002./(fi-24)+2.3953e+002./(fi-25)-2.7219e+002./(fi-26)+3.1937e+002./(fi-27)-3.8469e+002./(fi-28)+4.1010e+002./(fi-29)-4.0647e+002./(fi-30)+3.9195e+002./(fi-31)-3.8469e+002./(fi-32)+3.7381e+002./(fi-33)-3.5566e+002./(fi-34)+3.4114e+002./(fi-35)-3.1574e+002./(fi-36)+3.1211e+002./(fi-37)-3.0485e+002./(fi-38)+3.0122e+002./(fi-39)-2.8671e+002./(fi-40)+2.7945e+002./(fi-41)-2.6130e+002./(fi-42)+2.5404e+002./(fi-43)-2.4679e+002./(fi-44)+2.4316e+002./(fi-45)-2.4316e+002./(fi-46)+1.2702e+002./(fi-47)-1.2702e+002./ffi)./(3.6292*(1./(fi-1)-1./(fi-2)+1./(fi-3)-1./(fi-4)+1./(fi-5)-1./(fi-6)+1./(fi-7)-1./(fi-8)+1./(fi-9)-1./(fi-10)+1./(fi-11)-1./(fi-12)+1./(fi-13)-1./(fi-14)+1./(fi-15)-1./(fi-16)+1./(fi-17)-1./(fi-18)+1./(fi-19)-1./(fi-20)+1./(fi-21)-1./(fi-22)+1./(fi-23)-1./(fi-24)+1./(fi-25)-1./(fi-26)+1./(fi-27)-1./(fi-28)+1./(fi-29)-1./(fi-30)+1./(fi-31)-1./(fi-32)+1./(fi-33)-1./(fi-34)+1./(fi-35)-1./(fi-36)+1./(fi-37)-1./(fi-38)+1./(fi-39)-1./(fi-40)+1./(fi-41)-1./(fi-42)+1./(fi-43)-1./(fi-44)+1./(fi-45)-1./(fi-46)+0.5./(fi-47)-0.5./ffi)); (8-16)

$VS_{BLI}^{l\ W_4}(f_x)$=(2.6493e+002./(fi-1)-2.8308e+002./(fi-2)+2.8671e+002./(fi-3)-3.4477e+002./(fi-4)+3.5929e+002./(fi-5)-3.5566e+002./(fi-6)+3.5566e+002./(fi-7)-3.7381e+002./(fi-8)+3.7018e+002./(fi-9)-3.5566e+002./(fi-10)+3.4840e+002./(fi-11)-3.4840e+002./(fi-12)+3.5203e+002./(fi-13)-3.5203e+002./(fi-14)+3.4840e+002./(fi-15)-3.4114e+002./(fi-16)+3.4477e+002./(fi-17)-3.5566e+002./(fi-18)+3.6292e+002./(fi-19)-3.5929e+002./(fi-20)+3.7744e+002./(fi-21)-3.8107e+002./(fi-22)+3.3389e+002./(fi-23)-2.4679e+002./(fi-24)+2.6130e+002./(fi-25)-2.6493e+002./(fi-26)+2.7219e+002./(fi-27)-3.2663e+002./(fi-28)+3.5203e+002./(fi-29)-3.8107e+002./(fi-30)+3.7744e+002./(fi-31)-3.7018e+002./(fi-32)+3.5566e+002./(fi-33)-3.2300e+002./(fi-34)+3.0848e+002./(fi-35)-3.0485e+002./(fi-36)+3.0122e+002./(fi-37)-2.9396e+002./(fi-38)+2.7582e+002./(fi-39)-2.7219e+002./(fi-40)+2.6856e+002./(fi-41)-2.6493e+002./(fi-42)+2.6493e+002./(fi-43)-2.6130e+002./(fi-44)+2.5041e+002./(fi-45)-2.4316e+002./(fi-46)+1.2702e+002./(fi-47)-1.2702e+002./ffi)./(3.6292*(1./(fi-1)-1./(fi-2)+1./(fi-3)-1./(fi-4)+1./(fi-5)-1./(fi-6)+1./(fi-7)-1./(fi-8)+1./(fi-9)-1./(fi-10)+1./(fi-11)-1./(fi-12)+1./(fi-13)-1./(fi-14)+1./(fi-15)-1./(fi-16)+1./(fi-17)-1./(fi-18)+1./(fi-19)-1./(fi-20)+1./(fi-21)-1./(fi-22)+1./(fi-23)-1./(fi-24)+1./(fi-25)-1./(fi-26)+1./(fi-27)-1./(fi-28)+1./(fi-29)-1./(fi-30)+1./(fi-31)-1./(fi-32)+1./(fi-33)-1./(fi-34)+1./(fi-35)-1./(fi-36)+1./(fi-37)-1./(fi-38)+1./(fi-39)-1./(fi-40)+1./(fi-41)-1./(fi-42)+1./(fi-43)-1./(fi-44)+1./(fi-45)-1./(fi-46)+0.5./(fi-47)-0.5./ffi)); (8-17)

$VS_{BLI}^{l\ W_5}(f_x)$=1./(1.9077*(1./(fi-1)-1./(fi-2)+1./(fi-3)-1./(fi-4)+1./(fi-5)-1./(fi-6)+1./(fi-7)-1./(fi-8)+1./(fi-9)-1./(fi-10)+1./(fi-11)-1./(fi-12)+1./(fi-13)-1./(fi-14)+1./(fi-15)-1./(fi-16)+1./(fi-17)-1./(fi-18)+1./(fi-19)-1./(fi-20)+1./(fi-21)-1./(fi-22)+1./(fi-23)-1./(fi-24)+1./(fi-25)-1./(fi-26)+1./(fi-27)-1./(fi-28)+1./(fi-29)-1./(fi-30)+1./(fi-31)-1./(fi-32)+1./(fi-33)-1./(fi-34)+1./(fi-35)-1./(fi-36)+1./(fi-37)-1./(fi-38)+1./(fi-39)-1./(fi-40)+1./(fi-41)-1./(fi-42)+1./(fi-43)-1./(fi-44)+1./(fi-45)-1./(fi-46)+1./(fi-47)-1./(fi-48)+1./(fi-49)-1./(fi-50)+1./(fi-51)-1./(fi-52)+1./(fi-53)-1./(fi-54)+1./(fi-55)-1./(fi-56)+1./(fi-57)-1./(fi-58)+0.5./(fi-59)-0.5./ffi)).*(1.3354e+002./(fi-1)-1.3735e+002./(fi-2)+1.3735e+002./(fi-3)-1.4689e+002./(fi-4)+1.5834e+002./(fi-5)-1.9077e+002./(fi-6)+1.9649e+002./(fi-7)-2.0030e+002./(fi-8)+2.1366e+002./(fi-9)-2.2701e+002./(fi-10)+2.2510e+002./(fi-11)-2.3273e+002./(fi-12)+2.2892e+002./(fi-13)-2.1557e+002./(fi-14)+2.0412e+002./(fi-15)-2.0221e+002./(fi-16)+2.0984e+002./(fi-17)-1.9840e+002./(fi-18)+1.8886e+002./(fi-19)-1.6597e+002./(fi-20)+1.6024e+002./(fi-21)-1.6406e+002./(fi-22)+1.6024e+002./(fi-23)-1.6978e+002./(fi-24)+1.6024e+002./(fi-25)-1.5834e+002./(fi-26)+1.6024e+002./(fi-27)-1.6978e+002./(fi-28)+1.8504e+002./(fi-29)-2.0603e+002./(fi-30)+1.9840e+002./(fi-31)-1.8695e+002./(fi-32)+1.9077e+002./(fi-33)-1.8695e+002./(fi-34)+1.9840e+002./(fi-35)-1.8123e+002./(fi-36)+1.7550e+002./(fi-37)-1.6978e+002./(fi-38)+1.7741e+002./(fi-39)-1.7741e+002./(fi-40)+1.7741e+002./(fi-41)-1.7741e+002./(fi-42)+1.8123e+002./(fi-43)-1.7932e+002./(fi-44)+1.7550e+002./(fi-45)-1.7550e+002./(fi-46)+1.7169e+002./(fi-47)-1.7360e+002./(fi-48)+1.7360e+002./(fi-49)-1.6597e+002./(fi-50)+1.6024e+002./(fi-51)-1.4689e+002./(fi-52)+1.4307e+002./(fi-53)-1.3354e+002./(fi-54)+1.3735e+002./(fi-55)-1.3163e+002./(fi-56)+1.2781e+002./(fi-57)-1.2972e+002./(fi-58)+6.6768e+001./(fi-59)-6.6768e+001./ffi); (8-18)

$VS_{BLI}^{l\ W_6}(f_x)$=(3.5501e+002./(fi-1)-3.6008e+002./(fi-2)+3.7530e+002./(fi-3)-3.6516e+002./(fi-4)+4.5137e+002./(fi-5)-5.5281e+002./(fi-6)+6.9481e+002./(fi-7)-6.6438e+002./(fi-8)+6.8974e+002./(fi-9)-6.8974e+002./(fi-10)+6.4410e+002./(fi-11)-5.9338e+002./(fi-12)+5.1730e+002./(fi-13)-4.7673e+002./(fi-14)+4.5645e+002./(fi-15)-4.4123e+002./(fi-16)+4.4123e+002./(fi-17)-4.6152e+002./(fi-18)+4.6659e+002./(fi-19)-5.0716e+002./(fi-20)+4.8687e+002./(fi-21)-4.6152e+002./(fi-22)+3.9559e+002./(fi-23)-3.9559e+002./(fi-24)+4.5137e+002./(fi-25)-5.4773e+002./(fi-26)+5.2745e+002./(fi-27)-5.2238e+002./(fi-28)+4.2602e+002./(fi-29)-3.8544e+002./(fi-30)+4.0066e+002./(fi-31)-4.4630e+002./(fi-32)+4.9195e+002./(fi-33)-4.9702e+002./(fi-34)+5.0716e+002./(fi-35)-5.2238e+002./(fi-36)+5.0716e+002./(fi-37)-4.8687e+002./(fi-38)+4.1587e+002./(fi-39)-3.9559e+002./(fi-40)+3.7530e+002./(fi-41)-3.6516e+002./(fi- (8-19)

42)+1.7751e+002./(fi-43)-1.7751e+002./ffi)./(5.0716*(1./(fi-1)-1./(fi-2)+1./(fi-3)-1./(fi-4)+1./(fi-5)-1./(fi-6)+1./(fi-7)-1./(fi-8)+1./(fi-9)-1./(fi-10)+1./(fi-11)-1./(fi-12)+1./(fi-13)-1./(fi-14)+1./(fi-15)-1./(fi-16)+1./(fi-17)-1./(fi-18)+1./(fi-19)-1./(fi-20)+1./(fi-21)-1./(fi-22)+1./(fi-23)-1./(fi-24)+1./(fi-25)-1./(fi-26)+1./(fi-27)-1./(fi-28)+1./(fi-29)-1./(fi-30)+1./(fi-31)-1./(fi-32)+1./(fi-33)-1./(fi-34)+1./(fi-35)-1./(fi-36)+1./(fi-37)-1./(fi-38)+1./(fi-39)-1./(fi-40)+1./(fi-41)-1./(fi-42)+0.5./(fi-43)-0.5./ffi));

$VS_{BLI}^{l\ W_7}(f_x)$=(6.0906e+002./(fi-2)-6.0048e+002./(fi-1)-6.6053e+002./(fi-3)+6.6911e+002./(fi-4)-7.2916e+002./(fi-5)+8.3210e+002./(fi-6)-8.6641e+002./(fi-7)+9.0930e+002./(fi-8)-9.0930e+002./(fi-9)+9.6935e+002./(fi-10)-1.0122e+003./(fi-11)+1.0466e+003./(fi-12)-1.0466e+003./(fi-13)+1.0122e+003./(fi-14)-9.8651e+002./(fi-15)+9.7793e+002./(fi-16)-9.1788e+002./(fi-17)+8.7499e+002./(fi-18)-8.6641e+002./(fi-19)+8.6641e+002./(fi-20)-8.7499e+002./(fi-21)+8.7499e+002./(fi-22)-9.1788e+002./(fi-23)+9.1788e+002./(fi-24)-8.8357e+002./(fi-25)+7.8063e+002./(fi-26)-6.6911e+002./(fi-27)+6.2622e+002./(fi-28)-6.6911e+002./(fi-29)+7.4631e+002./(fi-30)-8.4925e+002./(fi-31)+8.7499e+002./(fi-32)-8.8357e+002./(fi-33)+8.5783e+002./(fi-34)-8.4925e+002./(fi-35)+8.6641e+002./(fi-36)-8.2352e+002./(fi-37)+7.9778e+002./(fi-38)-7.2916e+002./(fi-39)+6.4337e+002./(fi-40)-5.9190e+002./(fi-41)+6.0906e+002./(fi-42)-6.1764e+002./(fi-43)+3.0024e+002./(fi-44)+3.0024e+002./ffi)./(8.5783*(1./(fi-2)-1./(fi-1)-1./(fi-3)+1./(fi-4)-1./(fi-5)+1./(fi-6)-1./(fi-7)+1./(fi-8)-1./(fi-9)+1./(fi-10)-1./(fi-11)+1./(fi-12)-1./(fi-13)+1./(fi-14)-1./(fi-15)+1./(fi-16)-1./(fi-17)+1./(fi-18)-1./(fi-19)+1./(fi-20)-1./(fi-21)+1./(fi-22)-1./(fi-23)+1./(fi-24)-1./(fi-25)+1./(fi-26)-1./(fi-27)+1./(fi-28)+1./(fi-29)-1./(fi-30)+1./(fi-31)-1./(fi-32)+1./(fi-33)-1./(fi-34)+1./(fi-35)-1./(fi-36)+1./(fi-37)-1./(fi-38)+1./(fi-39)-1./(fi-40)+1./(fi-41)-1./(fi-42)+1./(fi-43)-1./(fi-44)+1./(fi-45)-1./(fi-46)+1./(fi-47)-1./(fi-48)+1./(fi-49)-1./(fi-50)+0.5./(fi-51)-0.5./ffi)).*(1.2929e+002./(fi-1)-1.3113e+002./(fi-2)+1.3852e+002./(fi-3)-1.5330e+002./(fi-4)+1.6622e+002./(fi-5)-1.8285e+002./(fi-6)+2.0686e+002./(fi-7)-2.0316e+002./(fi-8)+2.1978e+002./(fi-9)-2.1609e+002./(fi-10)+2.1424e+002./(fi-11)-2.0686e+002./(fi-12)+1.9762e+002./(fi-13)-1.9393e+002./(fi-14)+1.8654e+002./(fi-15)-1.8654e+002./(fi-16)+1.7915e+002./(fi-17)-1.6807e+002./(fi-18)+1.5884e+002./(fi-19)-1.5699e+002./(fi-20)+1.6438e+002./(fi-21)-1.6622e+002./(fi-22)+1.8100e+002./(fi-23)-1.8839e+002./(fi-24)+1.8285e+002./(fi-25)-1.6622e+002./(fi-26)+1.5330e+002./(fi-27)-1.5145e+002./(fi-28)+1.5884e+002./(fi-29)-1.6253e+002./(fi-30)+1.9023e+002./(fi-31)-2.0316e+002./(fi-32)+2.0316e+002./(fi-33)-1.9577e+002./(fi-34)+1.9208e+002./(fi-35)-1.8654e+002./(fi-36)+1.9577e+002./(fi-37)-1.9947e+002./(fi-38)+1.9762e+002./(fi-39)-1.8654e+002./(fi-40)+1.8839e+002./(fi-41)-1.7915e+002./(fi-42)+1.6807e+002./(fi-43)-1.6807e+002./(fi-44)+1.6253e+002./(fi-45)-1.5514e+002./(fi-46)+1.4591e+002./(fi-47)-1.4221e+002./(fi-48)+1.3667e+002./(fi-49)-1.3483e+002./(fi-50)+6.4643e+001./(fi-51)-6.4643e+001./ffi); (8-20)

$VS_{BLI}^{l\ W_8}(f_x)$=1./(1.8469*(1./(fi-1)-1./(fi-2)+1./(fi-3)-1./(fi-4)+1./(fi-5)-1./(fi-6)+1./(fi-7)-1./(fi-8)+1./(fi-9)-1./(fi-10)+1./(fi-11)-1./(fi-12)+1./(fi-13)-1./(fi-14)+1./(fi-15)-1./(fi-16)+1./(fi-17)-1./(fi-18)+1./(fi-19)-1./(fi-20)+1./(fi-21)-1./(fi-22)+1./(fi-23)-1./(fi-24)+1./(fi-25)-1./(fi-26)+1./(fi-27)-1./(fi-28)+1./(fi-29)-1./(fi-30)+1./(fi-31)-1./(fi-32)+1./(fi-33)-1./(fi-34)+1./(fi-35)-1./(fi-36)+1./(fi-37)-1./(fi-38)+1./(fi-39)-1./(fi-40)+1./(fi-41)-1./(fi-42)+1./(fi-43)-1./(fi-44)+1./(fi-45)-1./(fi-46)+1./(fi-47)-1./(fi-48)+1./(fi-49)-1./(fi-50)+0.5./(fi-51)-0.5./ffi)).*(1.2929e+002./(fi-1)-1.3113e+002./(fi-2)+1.3852e+002./(fi-3)-1.5330e+002./(fi-4)+1.6622e+002./(fi-5)-1.8285e+002./(fi-6)+2.0686e+002./(fi-7)-2.0316e+002./(fi-8)+2.1978e+002./(fi-9)-2.1609e+002./(fi-10)+2.1424e+002./(fi-11)-2.0686e+002./(fi-12)+1.9762e+002./(fi-13)-1.9393e+002./(fi-14)+1.8654e+002./(fi-15)-1.8654e+002./(fi-16)+1.7915e+002./(fi-17)-1.6807e+002./(fi-18)+1.5884e+002./(fi-19)-1.5699e+002./(fi-20)+1.6438e+002./(fi-21)-1.6622e+002./(fi-22)+1.8100e+002./(fi-23)-1.8839e+002./(fi-24)+1.8285e+002./(fi-25)-1.6622e+002./(fi-26)+1.5330e+002./(fi-27)-1.5145e+002./(fi-28)+1.5884e+002./(fi-29)-1.6253e+002./(fi-30)+1.9023e+002./(fi-31)-2.0316e+002./(fi-32)+2.0316e+002./(fi-33)-1.9577e+002./(fi-34)+1.9208e+002./(fi-35)-1.8654e+002./(fi-36)+1.9577e+002./(fi-37)-1.9947e+002./(fi-38)+1.9762e+002./(fi-39)-1.8654e+002./(fi-40)+1.8839e+002./(fi-41)-1.7915e+002./(fi-42)+1.6807e+002./(fi-43)-1.6807e+002./(fi-44)+1.6253e+002./(fi-45)-1.5514e+002./(fi-46)+1.4591e+002./(fi-47)-1.4221e+002./(fi-48)+1.3667e+002./(fi-49)-1.3483e+002./(fi-50)+6.4643e+001./(fi-51)-6.4643e+001./ffi); (8-21)

$VS_{BLI}^{l\ W_9}(f_x)$=1./(1.8469*(1./(fi-1)-1./(fi-2)+1./(fi-3)-1./(fi-4)+1./(fi-5)-1./(fi-6)+1./(fi-7)-1./(fi-8)+1./(fi-9)-1./(fi-10)+1./(fi-11)-1./(fi-12)+1./(fi-13)-1./(fi-14)+1./(fi-15)-1./(fi-16)+1./(fi-17)-1./(fi-18)+1./(fi-19)-1./(fi-20)+1./(fi-21)-1./(fi-22)+1./(fi-23)-1./(fi-24)+1./(fi-25)-1./(fi-26)+1./(fi-27)-1./(fi-28)+1./(fi-29)-1./(fi-30)+1./(fi-31)-1./(fi-32)+1./(fi-33)-1./(fi-34)+1./(fi-35)-1./(fi-36)+1./(fi-37)-1./(fi-38)+1./(fi-39)-1./(fi-40)+1./(fi-41)-1./(fi-42)+1./(fi-43)-1./(fi-44)+1./(fi-45)-1./(fi-46)+1./(fi-47)-1./(fi-48)+1./(fi-49)-1./(fi-50)+0.5./(fi-51)-0.5./ffi)).*(1.2190e+002./(fi-1)-1.2744e+002./(fi-2)+1.2744e+002./(fi-3)-1.6068e+002./(fi-4)+1.7176e+002./(fi-5)-1.6807e+002./(fi-6)+1.6622e+002./(fi-7)-1.6438e+002./(fi-8)+1.5884e+002./(fi-9)-1.6253e+002./(fi-10)+1.5884e+002./(fi-11)-1.5514e+002./(fi-12)+1.5514e+002./(fi-13)-1.5330e+002./(fi-14)+1.5514e+002./(fi-15)-1.6068e+002./(fi-16)+1.6807e+002./(fi-17)-1.4775e+002./(fi-18)+1.3483e+002./(fi-19)-1.3667e+002./(fi-20)+1.5145e+002./(fi-21)-1.6622e+002./(fi-22)+1.6992e+002./(fi-23)-1.7176e+002./(fi-24)+1.7176e+002./(fi-25)-1.6438e+002./(fi-26)+1.6253e+002./(fi-27)-1.4406e+002./(fi-28)+1.4406e+002./(fi-29)-1.4037e+002./(fi-30)+1.3113e+002./(fi-31)-1.4221e+002./(fi-32)+1.3667e+002./(fi-33)-1.3852e+002./(fi-34)+1.3667e+002./(fi-35)-1.3667e+002./(fi-36)+1.2929e+002./(fi-37)-1.2374e+002./(fi-38)+1.2190e+002./(fi-39)-1.2190e+002./(fi-40)+1.1820e+002./(fi-41)-1.2374e+002./(fi-42)+1.2190e+002./(fi-43)-1.1820e+002./(fi-44)+1.2005e+002./(fi-45)-1.1636e+002./(fi-46)+1.1636e+002./(fi-47)-1.1266e+002./(fi-48)+1.1636e+002./(fi-49)-1.2190e+002./(fi-50)+6.4643e+001./(fi-51)-6.4643e+001./ffi); (8-22)

$VS_{BLI}^{l\ W_{10}}(f_x)$=1./(3.6980*(1./(fi-1)-1./(fi-2)+1./(fi-3)-1./(fi-4)+1./(fi-5)-1./(fi-6)+1./(fi-7)-1./(fi-8)+1./(fi-9)-1./(fi-10)+1./(fi-11)-1./(fi-12)+1./(fi-13)-1./(fi-14)+1./(fi-15)-1./(fi-16)+1./(fi-17)-1./(fi-18)+1./(fi-19)-1./(fi-20)+1./(fi-21)-1./(fi-22)+1./(fi-23)-1./(fi-24)+1./(fi-25)-1./(fi-26)+1./(fi-27)-1./(fi-28)+1./(fi-29)-1./(fi-30)+1./(fi-31)-1./(fi-32)+1./(fi-33)-1./(fi-34)+1./(fi-35)-1./(fi-36)+1./(fi-37)-1./(fi-38)+1./(fi-39)-1./(fi-40)+1./(fi-41)-1./(fi-42)+1./(fi-43)-1./(fi-44)+1./(fi-45)-1./(fi-46)+1./(fi-47)-1./(fi-48)+1./(fi-49)-1./(fi-50)+1./(fi-51)-1./(fi-52)+0.5./(fi-53)-0.5./ffi)).*(2.5886e+002./(fi-1)-2.6256e+002./(fi-2)+2.7365e+002./(fi-3)-3.0324e+002./(fi-4)+3.2173e+002./(fi-5)-3.1803e+002./(fi-6)+3.3282e+002./(fi-7)-3.7720e+002./(fi-8)+4.0678e+002./(fi-9)-4.2527e+002./(fi-10)+4.2527e+002./(fi-11)-4.1418e+002./(fi-12)+3.9569e+002./(fi-13)-3.7350e+002./(fi-14)+3.7350e+002./(fi-15)-3.5131e+002./(fi-16)+3.4392e+002./(fi-17)-3.4022e+002./(fi-18)+3.3652e+002./(fi-19)-3.4392e+002./(fi-20)+3.5871e+002./(fi-21)-3.7350e+002./(fi-22)+3.6980e+002./(fi-23)-3.4022e+002./(fi-24)+3.1803e+002./(fi-25)-2.9954e+002./(fi-26)+3.0324e+002./(fi-27)-3.4022e+002./(fi-28)+3.6241e+002./(fi-29)-3.9569e+002./(fi-30)+4.2897e+002./(fi-31)-4.5486e+002./(fi-32)+4.6965e+002./(fi-33)-4.7704e+002./(fi-34)+4.8814e+002./(fi-35)-4.6965e+002./(fi-36)+4.5855e+002./(fi-37)-4.4376e+002./(fi-38)+4.0308e+002./(fi-39)-3.6610e+002./(fi-40)+3.5131e+002./(fi-41)-3.4392e+002./(fi-42)+3.3282e+002./(fi-43)-3.2543e+002./(fi-44)+3.1433e+002./(fi-45)-3.1063e+002./(fi-46)+2.9954e+002./(fi-47)-2.8475e+002./(fi-48)+2.7735e+002./(fi-49)-2.7735e+002./(fi-50)+2.6626e+002./(fi-51)-2.5886e+002./(fi-52)+1.2943e+002./(fi-53)-1.2943e+002./ffi); (8-23)

$VS_{BLI}^{l\ W_{11}}(f_x)$=(5.8845e+002./(fi-1)-5.7992e+002./(fi-2)+6.0551e+002./(fi-3)-6.3962e+002./(fi-4)+6.4815e+002./(fi-5)-7.5049e+002./(fi-6)+7.9313e+002./(fi-7)-7.7607e+002./(fi-8)+8.0166e+002./(fi-9)-7.9313e+002./(fi-10)+8.0166e+002./(fi-11)-7.9313e+002./(fi-12)+7.9313e+002./(fi-13)-7.9313e+002./(fi-14)+8.1019e+002./(fi-15)-8.3577e+002./(fi-16)+8.0166e+002./(fi-17)-7.7607e+002./(fi-18)+8.6136e+002./(fi-19)-8.6136e+002./(fi-20)+6.9079e+002./(fi-21)-7.2490e+002./(fi-22)+6.7373e+002./(fi-23)-6.3109e+002./(fi-24)+6.3962e+002./(fi-25)-6.8226e+002./(fi-26)+7.7607e+002./(fi-27)-7.9313e+002./(fi-28)+9.9781e+002./(fi-29)-8.9547e+002./(fi-30)+8.1872e+002./(fi-31)-8.2724e+002./(fi-32)+8.0166e+002./(fi-33)-8.1019e+002./(fi-34)+7.9313e+002./(fi-35)-7.5902e+002./(fi-36)+7.0785e+002./(fi-37)-6.9079e+002./(fi-38)+6.9932e+002./(fi-39)-6.9932e+002./(fi-40)+6.9932e+002./(fi-41)-6.7373e+002./(fi-42)+6.3962e+002./(fi-43)-6.4815e+002./(fi-44)+6.4815e+002./(fi-45)-6.9079e+002./(fi-46)+6.3109e+002./(fi-47)-6.4815e+002./(fi-48)+2.9849e+002./(fi-49)-2.9849e+002./ffi)./(8.5283*(1./(fi-1)-1./(fi-2)+1./(fi-3)-1./(fi-4)+1./(fi-5)-1./(fi-6)+1./(fi-7)-1./(fi-8)+1./(fi-9)-1./(fi-10)+1./(fi-11)-1./(fi-12)+1./(fi-13)-1./(fi-14)+1./(fi-15)-1./(fi-16)+1./(fi-17)-1./(fi-18)+1./(fi-19)-1./(fi-20)+1./(fi-21)-1./(fi-22)+1./(fi-23)-1./(fi-24)+1./(fi-25)-1./(fi-26)+1./(fi-27)-1./(fi-28)+1./(fi-29)-1./(fi-30)+1./(fi-31)-1./(fi-32)+1./(fi-33)-1./(fi-34)+1./(fi-35)-1./(fi-36)+1./(fi-37)-1./(fi-38)+1./(fi-39)-1./(fi-40)+1./(fi-41)-1./(fi-42)+1./(fi-43)-1./(fi-44)+1./(fi-45)-1./(fi-46)+1./(fi-47)-1./(fi-48)+0.5./(fi-49)-0.5./ffi)); (8-24)

$VS_{BLI}^{l\ W_{12}}(f_x)$=1./(9.0949*(1./(fi-2)-1./(fi-1)-1./(fi-3)+1./(fi-4)-1./(fi-5)+1./(fi-6)-1./(fi-7)+1./(fi-8)-1./(fi-9)+1./(fi-10)-1./(fi-11)+1./(fi-12)-1./(fi-13)+1./(fi-14)-1./(fi-15)+1./(fi-16)-1./(fi-17)+1./(fi-18)-1./(fi-19)+1./(fi-20)-1./(fi-21)+1./(fi-22)-1./(fi-23)+1./(fi-24)-1./(fi-25)+1./(fi-26)-1./(fi-27)+1./(fi-28)-1./(fi-29)+1./(fi-30)-1./(fi-31)+1./(fi-32)-1./(fi-33)+1./(fi-34)-1./(fi-35)+1./(fi-36)-1./(fi-37)+1./(fi-38)-1./(fi-39)+0.5./(fi-40)+0.5./ffi)).*(6.5484e+002./(fi-2)-6.4574e+002./(fi-1)-6.6393e+002./(fi-3)+6.8212e+002./(fi-4)-7.4579e+002./(fi-5)+7.9126e+002./(fi-6)-8.5493e+002./(fi-7)+8.8221e+002./(fi-8)-8.7311e+002./(fi-9)+8.9130e+002./(fi-10)-9.0949e+002./(fi-11)+9.0040e+002./(fi-12)-8.7311e+002./(fi-13)+8.6402e+002./(fi-14)-8.6402e+002./(fi-15)+8.7311e+002./(fi-16)-8.8221e+002./(fi-17)+9.3678e+002./(fi-18)-9.6406e+002./(fi-19)+8.6402e+002./(fi-20)-8.1855e+002./(fi-21)+6.9122e+002./(fi-22)-7.2760e+002./(fi-23)+8.0945e+002./(fi-24)-8.1855e+002./(fi-25)+8.6402e+002./(fi-26)-8.6402e+002./(fi-27)+8.1855e+002./(fi-28)-8.0036e+002./(fi-29)+7.7307e+002./(fi-30)-7.3669e+002./(fi-31)+7.4579e+002./(fi-32)-7.2760e+002./(fi-33)+7.2760e+002./(fi-34)-7.2760e+002./(fi-35)+7.1850e+002./(fi-36)-7.0031e+002./(fi-37)+6.8212e+002./(fi-38)-6.4574e+002./(fi-39)+3.1832e+002./(fi-40)+3.1832e+002./ffi); (8-25)

$VS_{BLI}^{l\ W_{13}}(f_x)$=(3.4994e+002./(fi-1)-3.9559e+002./(fi-2)+4.6152e+002./(fi-3)-4.8687e+002./(fi-4)+5.1730e+002./(fi-5)-5.4266e+002./(fi-6)+5.6295e+002./(fi-7)-5.7816e+002./(fi-8)+5.6802e+002./(fi-9)-5.6295e+002./(fi-10)+5.5788e+002./(fi-11)-5.6295e+002./(fi-12)+5.5281e+002./(fi-13)-5.3759e+002./(fi-14)+5.4266e+002./(fi-15)-4.6659e+002./(fi-16)+3.9051e+002./(fi-17)-4.1080e+002./(fi-18)+3.9559e+002./(fi-19)-4.6152e+002./(fi-20)+5.4266e+002./(fi-21)-5.3759e+002./(fi-22)+4.9195e+002./(fi-23)-4.7673e+002./(fi-24)+4.8180e+002./(fi-25)-4.6659e+002./(fi-26)+4.5645e+002./(fi-27)-4.4123e+002./(fi-28)+4.3616e+002./(fi-29)-4.3616e+002./(fi-30)+4.3109e+002./(fi-31)-4.0573e+002./(fi-32)+4.2094e+002./(fi-33)-4.1080e+002./(fi-34)+3.9051e+002./(fi-35)-4.1080e+002./(fi-36)+4.1080e+002./(fi-37)-4.0066e+002./(fi-38)+3.9051e+002./(fi-39)-3.7530e+002./(fi-40)+3.6516e+002./(fi-41)-3.5501e+002./(fi-42)+1.7751e+002./(fi-43)-1.7751e+002./ffi)./(5.0716*(1./(fi-1)-1./(fi-2)+1./(fi-3)-1./(fi-4)+1./(fi-5)-1./(fi-6)+1./(fi-7)-1./(fi-8)+1./(fi-9)-1./(fi-10)+1./(fi-11)-1./(fi-12)+1./(fi-13)-1./(fi-14)+1./(fi-15)-1./(fi-16)+1./(fi-17)-1./(fi-18)+1./(fi-19)-1./(fi-20)+1./(fi-21)-1./(fi-22)+1./(fi-23)-1./(fi-24)+1./(fi-25)-1./(fi-26)+1./(fi-27)-1./(fi-28)+1./(fi-29)-1./(fi-30)+1./(fi-31)-1./(fi-32)+1./(fi-33)-1./(fi-34)+1./(fi-35)-1./(fi-36)+1./(fi-37)-1./(fi-38)+1./(fi-39)-1./(fi-40)+1./(fi-41)-1./(fi-42)+0.5./(fi-43)-0.5./ffi)); (8-26)

c. CORNER VISUAL SPEECH SIGNALS

$VS_{BLI}^{c\ W_1}(f_x)$=(6.4715e+002./(fi-1)-6.2866e+002./(fi-2)+6.2866e+002./(fi-3)-6.2866e+002./(fi-4)+6.2866e+002./(fi-5)-6.3976e+002./(fi-6)+6.6194e+002./(fi-7)-6.5455e+002./(fi-8)+6.5825e+002./(fi-9)-6.6194e+002./(fi-10)+6.6564e+002./(fi-11)-6.8043e+002./(fi-12)+6.6934e+002./(fi-13)-6.7674e+002./(fi-14)+6.6934e+002./(fi-15)-6.6934e+002./(fi-16)+6.6564e+002./(fi-17)-6.6194e+002./(fi-18)+6.6564e+002./(fi-19)-6.6934e+002./(fi-20)+6.6564e+002./(fi-21)-6.6934e+002./(fi-22)+6.6194e+002./(fi-23)-6.3236e+002./(fi-24)+6.1017e+002./(fi-25)-6.1017e+002./(fi-26)+6.0647e+002./(fi-27)-5.9168e+002./(fi-28)+5.9908e+002./(fi-29)-6.0647e+002./(fi-30)+6.1757e+002./(fi-31)-6.1017e+002./(fi-32)+6.2866e+002./(fi-33)-6.2866e+002./(fi-34)+6.2866e+002./(fi-35)-6.4345e+002./(fi-36)+6.4345e+002./(fi-37)-6.5085e+002./(fi-38)+6.4715e+002./(fi-39)-6.3976e+002./(fi-40)+6.3976e+002./(fi-41)-6.4715e+002./(fi-42)+6.3606e+002./(fi-43)-6.3606e+002./(fi-44)+6.2496e+002./(fi-45)-6.3236e+002./(fi-46)+6.4345e+002./(fi-47)-6.3976e+002./(fi-48)+6.4345e+002./(fi-49)-6.2127e+002./(fi-50)+6.3976e+002./(fi-51)-6.3606e+002./(fi-52)+3.2173e+002./(fi-53)-3.2173e+002./ffi)./(3.6980*(1./(fi-1)-1./(fi-2)+1./(fi-3)-1./(fi-4)+1./(fi-5)-1./(fi-6)+1./(fi-7)-1./(fi-8)+1./(fi-9)-1./(fi-10)+1./(fi-11)-1./(fi-12)+1./(fi-13)-1./(fi-14)+1./(fi-15)-1./(fi-16)+1./(fi-17)-1./(fi-18)+1./(fi-19)-1./(fi-20)+1./(fi-21)-1./(fi-22)+1./(fi-23)-1./(fi-24)+1./(fi-25)-1./(fi-26)+1./(fi-27)-1./(fi-28)+1./(fi-29)-1./(fi-30)+1./(fi-31)-1./(fi-32)+1./(fi-33)-1./(fi-34)+1./(fi-35)-1./(fi-36)+1./(fi-37)-1./(fi-38)+1./(fi-39)-1./(fi-40)+1./(fi-41)-1./(fi-42)+1./(fi-43)-1./(fi-44)+1./(fi-45)-1./(fi-46)+1./(fi-47)-1./(fi-48)+1./(fi-49)-1./(fi-50)+1./(fi-51)-1./(fi-52)+0.5./(fi-53)-0.5./ffi)); (8-27)

$VS_{BLI}^{c\ W_2}(f_x)$=1./(8.5283*(1./(fi-1)-1./(fi-2)+1./(fi-3)-1./(fi-4)+1./(fi-5)-1./(fi-6)+1./(fi-7)-1./(fi-8)+1./(fi-9)-1./(fi-10)+1./(fi-11)-1./(fi-12)+1./(fi-13)-1./(fi-14)+1./(fi-15)-1./(fi-16)+1./(fi-17)-1./(fi-18)+1./(fi-19)-1./(fi-20)+1./(fi-21)-1./(fi-22)+1./(fi-23)-1./(fi-24)+1./(fi-25)-1./(fi-26)+1./(fi-27)-1./(fi-28)+1./(fi-29)-1./(fi-30)+1./(fi-31)-1./(fi-32)+1./(fi-33)-1./(fi-34)+1./(fi-35)-1./(fi-36)+1./(fi-37)-1./(fi-38)+1./(fi-39)-1./(fi-40)+1./(fi-41)-1./(fi-42)+1./(fi-43)-1./(fi-44)+1./(fi-45)-1./(fi-46)+1./(fi-47)-1./(fi-48)+0.5./(fi-49)-0.5./ffi)).*(1.4925e+003./(fi-1)-1.5010e+003./(fi-2)+1.5010e+003./(fi-3)-1.5095e+003./(fi-4)+1.5180e+003./(fi-5)-1.5351e+003./(fi-6)+1.5266e+003./(fi-7)-1.5436e+003./(fi-8)+1.5436e+003./(fi-9)-1.5436e+003./(fi-10)+1.5521e+003./(fi-11)-1.5607e+003./(fi-12)+1.5777e+003./(fi-13)-1.5863e+003./(fi-14)+1.5777e+003./(fi-15)-1.5692e+003./(fi-16)+1.5607e+003./(fi-17)-1.5521e+003./(fi-18)+1.5521e+003./(fi-19)-1.5436e+003./(fi-20)+1.5351e+003./(fi-21)-1.5351e+003./(fi-22)+1.5095e+003./(fi-23)-1.5010e+003./(fi-24)+1.4669e+003./(fi-25)-1.4669e+003./(fi-26)+1.4754e+003./(fi-27)-1.4498e+003./(fi-28)+1.4669e+003./(fi-29)-1.4413e+003./(fi-30)+1.4754e+003./(fi-31)-1.4669e+003./(fi-32)+1.4925e+003./(fi-33)-1.5095e+003./(fi-34)+1.5095e+003./(fi-35)-1.5095e+003./(fi-36)+1.5351e+003./(fi-37)-1.5180e+003./(fi-38)+1.5010e+003./(fi-39)-1.5010e+003./(fi-40)+1.5095e+003./(fi-41)-1.5351e+003./(fi-42)+1.5095e+003./(fi-43)-1.5266e+003./(fi-44)+1.5180e+003./(fi-45)-1.5351e+003./(fi-46)+1.5180e+003./(fi-47)-1.5180e+003./(fi-48)+7.4196e+002./(fi-49)-7.4196e+002./ffi); (8-28)

$VS_{BLI}^{c\ W_3}(f_x)$=(6.2422e+002./(fi-1)-6.1696e+002./(fi-2)+6.0608e+002./(fi-3)-6.1333e+002./(fi-4)+6.0970e+002./(fi-5)-6.1333e+002./(fi-6)+6.3511e+002./(fi-7)-6.2785e+002./(fi-8)+6.3511e+002./(fi-9)-6.3511e+002./(fi-10)+6.4963e+002./(fi-11)-6.5326e+002./(fi-12)+6.4600e+002./(fi-13)-6.4963e+002./(fi-14)+6.4237e+002./(fi-15)-6.4237e+002./(fi-16)+6.2059e+002./(fi-17)-6.0245e+002./(fi-18)+5.8793e+002./(fi-19)-5.8430e+002./(fi-20)+5.8430e+002./(fi-21)-5.8067e+002./(fi-22)+5.8430e+002./(fi-23)-5.8793e+002./(fi-24)+5.7341e+002./(fi-25)-5.9519e+002./(fi-26)+6.1333e+002./(fi-27)-6.2422e+002./(fi-28)+6.4600e+002./(fi-29)-6.4237e+002./(fi-30)+6.4237e+002./(fi-31)-6.4237e+002./(fi-32)+6.3874e+002./(fi-33)-6.3874e+002./(fi-34)+6.2785e+002./(fi-35)-6.3511e+002./(fi-36)+6.3874e+002./(fi-37)-6.2785e+002./(fi-38)+6.3874e+002./(fi-39)-6.2422e+002./(fi-40)+6.3511e+002./(fi-41)-6.4237e+002./(fi-42)+6.2785e+002./(fi-43)-6.3511e+002./(fi-44)+6.3148e+002./(fi-45)-6.4237e+002./(fi-46)+3.1574e+002./(fi-47)-3.1574e+002./ffi)./(3.6292*(1./(fi-1)-1./(fi-2)+1./(fi-3)-1./(fi-4)+1./(fi-5)-1./(fi-6)+1./(fi-7)-1./(fi-8)+1./(fi-9)-1./(fi-10)+1./(fi-11)-1./(fi-12)+1./(fi-13)-1./(fi-14)+1./(fi-15)-1./(fi-16)+1./(fi-17)-1./(fi-18)+1./(fi-19)-1./(fi-20)+1./(fi-21)-1./(fi-22)+1./(fi-23)-1./(fi-24)+1./(fi-25)-1./(fi-26)+1./(fi-27)-1./(fi-28)+1./(fi-29)-1./(fi-30)+1./(fi-31)-1./(fi-32)+1./(fi-33)-1./(fi-34)+1./(fi-35)-1./(fi-36)+1./(fi-37)-1./(fi-38)+1./(fi-39)-1./(fi-40)+1./(fi-41)-1./(fi-42)+1./(fi-43)-1./(fi-44)+1./(fi-45)-1./(fi-46)+0.5./(fi-47)-0.5./ffi)); (8-29)

$VS_{BLI}^{c\ W_4}(f_x)$=(6.4963e+002./(fi-1)-6.3148e+002./(fi-2)+6.3874e+002./(fi-3)-6.2059e+002./(fi-4)+6.2785e+002./(fi-5)-6.3874e+002./(fi-6)+6.4237e+002./(fi-7)-6.6051e+002./(fi-8)+6.6777e+002./(fi-9)-6.7503e+002./(fi-10)+6.6777e+002./(fi-11)-6.6414e+002./(fi-12)+6.6777e+002./(fi-13)-6.7503e+002./(fi-14)+6.7503e+002./(fi-15)-6.7503e+002./(fi-16)+6.6777e+002./(fi-17)-6.6414e+002./(fi-18)+6.6414e+002./(fi-19)-6.4237e+002./(fi-20)+6.3874e+002./(fi-21)-6.3148e+002./(fi-22)+6.2422e+002./(fi-23)-6.1333e+002./(fi-24)+6.0608e+002./(fi-25)-5.8793e+002./(fi-26)+5.9882e+002./(fi-27)-6.1333e+002./(fi-28)+6.3874e+002./(fi-29)-6.6051e+002./(fi-30)+6.5688e+002./(fi-31)-6.7140e+002./(fi-32)+6.6777e+002./(fi-33)-6.6051e+002./(fi-34)+6.6051e+002./(fi-35)-6.5688e+002./(fi-36)+6.6051e+002./(fi-37)-6.6051e+002./(fi-38)+6.5326e+002./(fi-39)-6.5688e+002./(fi-40)+6.4600e+002./(fi-41)-6.4600e+002./(fi-42)+6.4600e+002./(fi-43)-6.3874e+002./(fi-44)+6.4237e+002./(fi-45)-6.3511e+002./(fi-46)+3.1574e+002./(fi-47)-3.1574e+002./ffi)./(3.6292*(1./(fi-1)-1./(fi-2)+1./(fi-3)-1./(fi-4)+1./(fi-5)-1./(fi-6)+1./(fi-7)-1./(fi-8)+1./(fi-9)-1./(fi-10)+1./(fi-11)-1./(fi-12)+1./(fi-13)-1./(fi-14)+1./(fi-15)-1./(fi-16)+1./(fi-17)-1./(fi-18)+1./(fi-19)-1./(fi-20)+1./(fi-21)-1./(fi-22)+1./(fi-23)-1./(fi-24)+1./(fi-25)-1./(fi-26)+1./(fi-27)-1./(fi-28)+1./(fi-29)-1./(fi-30)+1./(fi-31)-1./(fi-32)+1./(fi-33)-1./(fi-34)+1./(fi-35)-1./(fi-36)+1./(fi-37)-1./(fi-38)+1./(fi-39)-1./(fi-40)+1./(fi-41)-1./(fi-42)+1./(fi-43)-1./(fi-44)+1./(fi-45)-1./(fi-46)+0.5./(fi-47)-0.5./ffi)); (8-30)

$VS_{BLI}^{c\ W_5}(f_x)$=(3.3384e+002./(fi-1)-3.3193e+002./(fi-2)+3.3384e+002./(fi-3)-3.4147e+002./(fi-4)+3.3193e+002./(fi-5)-3.3956e+002./(fi-6)+3.4338e+002./(fi-7)-3.5482e+002./(fi-8)+3.6436e+002./(fi-9)-3.6246e+002./(fi-10)+3.6436e+002./(fi-11)-3.6436e+002./(fi-12)+3.7009e+002./(fi-13)-3.7009e+002./(fi-14)+3.7009e+002./(fi-15)-3.6627e+002./(fi-16)+3.6627e+002./(fi-17)-3.6436e+002./(fi-18)+3.5482e+002./(fi-19)-3.5101e+002./(fi-20)+3.4719e+002./(fi-21)-3.4529e+002./(fi-22)+3.4338e+002./(fi-23)-3.2239e+002./(fi-24)+3.1858e+002./(fi-25)-3.2812e+002./(fi-26)+3.3193e+002./(fi-27)-3.2621e+002./(fi-28)+3.4529e+002./(fi-29)-3.5482e+002./(fi-30)+3.5101e+002./(fi-31)-3.5673e+002./(fi-32)+3.6055e+002./(fi-33)-3.5101e+002./(fi-34)+3.6055e+002./(fi-35)-3.5482e+002./(fi-36)+3.5482e+002./(fi-37)-3.5292e+002./(fi-38)+3.5292e+002./(fi-39)-3.5101e+002./(fi-40)+3.5101e+002./(fi-41)-3.4147e+002./(fi-42)+3.4529e+002./(fi-43)-3.4529e+002./(fi-44)+3.4338e+002./(fi-45)-3.5101e+002./(fi-46)+3.5292e+002./(fi-47)-3.4529e+002./(fi-48)+3.4338e+002./(fi-49)-3.3384e+002./(fi-50)+3.4910e+002./(fi-51)-3.3956e+002./(fi-52)+3.3766e+002./(fi-53)-3.3384e+002./(fi-54)+3.3003e+002./(fi-55)-3.2621e+002./(fi-56)+3.3003e+002./(fi-57)-3.2812e+002./(fi-58)+1.6597e+002./(fi-59)-1.6597e+002./ffi)./(1.9077*(1./(fi-1)-1./(fi-2)+1./(fi-3)-1./(fi-4)+1./(fi-5)-1./(fi-6)+1./(fi-7)-1./(fi-8)+1./(fi-9)-1./(fi-10)+1./(fi-11)-1./(fi-12)+1./(fi-13)-1./(fi-14)+1./(fi-15)-1./(fi-16)+1./(fi-17)-1./(fi-18)+1./(fi-19)-1./(fi-20)+1./(fi-21)-1./(fi-22)+1./(fi-23)-1./(fi-24)+1./(fi-25)-1./(fi-26)+1./(fi-27)-1./(fi-28)+1./(fi-29)-1./(fi-30)+1./(fi-31)-1./(fi-32)+1./(fi-33)-1./(fi-34)+1./(fi-35)-1./(fi-36)+1./(fi-37)-1./(fi-38)+1./(fi-39)-1./(fi-40)+1./(fi-41)-1./(fi-42)+1./(fi-43)-1./(fi-44)+1./(fi-45)-1./(fi-46)+1./(fi-47)-1./(fi-48)+1./(fi-49)-1./(fi-50)+1./(fi-51)-1./(fi-52)+1./(fi-53)-1./(fi-54)+1./(fi-55)-1./(fi-56)+1./(fi-57)-1./(fi-58)+0.5./(fi-59)-0.5./ffi)); (8-31)

$VS_{BLI}^{c}{}^{W_6}(f_x)$=(8.9260e+002./(fi-1)-9.2303e+002./(fi-2)+8.9260e+002./(fi-3)-9.0275e+002./(fi-4)+8.9768e+002./(fi-5)-8.9260e+002./(fi-6)+1.0042e+003./(fi-7)-1.0346e+003./(fi-8)+1.0397e+003./(fi-9)-1.0448e+003./(fi-10)+1.0448e+003./(fi-11)-1.0093e+003./(fi-12)+1.0143e+003./(fi-13)-1.0093e+003./(fi-14)+9.8896e+002./(fi-15)-9.8896e+002./(fi-16)+9.8389e+002./(fi-17)-9.7882e+002./(fi-18)+9.0782e+002./(fi-19)-8.7739e+002./(fi-20)+8.7739e+002./(fi-21)-8.7232e+002./(fi-22)+8.3682e+002./(fi-23)-8.4696e+002./(fi-24)+8.9260e+002./(fi-25)-9.1796e+002./(fi-26)+9.5346e+002./(fi-27)-9.4332e+002./(fi-28)+9.6361e+002./(fi-29)-9.4332e+002./(fi-30)+9.7882e+002./(fi-31)-9.9911e+002./(fi-32)+9.9911e+002./(fi-33)-9.8896e+002./(fi-34)+9.8896e+002./(fi-35)-9.7375e+002./(fi-36)+9.6361e+002./(fi-37)-9.6868e+002./(fi-38)+9.2303e+002./(fi-39)-8.8753e+002./(fi-40)+8.7739e+002./(fi-41)-8.8246e+002./(fi-42)+4.4123e+002./(fi-43)-4.4123e+002./ffi)./(5.0716*(1./(fi-1)-1./(fi-2)+1./(fi-3)-1./(fi-4)+1./(fi-5)-1./(fi-6)+1./(fi-7)-1./(fi-8)+1./(fi-9)-1./(fi-10)+1./(fi-11)-1./(fi-12)+1./(fi-13)-1./(fi-14)+1./(fi-15)-1./(fi-16)+1./(fi-17)-1./(fi-18)+1./(fi-19)-1./(fi-20)+1./(fi-21)-1./(fi-22)+1./(fi-23)-1./(fi-24)+1./(fi-25)-1./(fi-26)+1./(fi-27)-1./(fi-28)+1./(fi-29)-1./(fi-30)+1./(fi-31)-1./(fi-32)+1./(fi-33)-1./(fi-34)+1./(fi-35)-1./(fi-36)+1./(fi-37)-1./(fi-38)+1./(fi-39)-1./(fi-40)+1./(fi-41)-1./(fi-42)+0.5./(fi-43)-0.5./ffi)); (8-32)

$VS_{BLI}^{c}{}^{W_7}(f_x)$=(1.4840e+003./(fi-2)-1.4669e+003./(fi-1)-1.5012e+003./(fi-3)+1.4583e+003./(fi-4)-1.5012e+003./(fi-5)+1.5269e+003./(fi-6)-1.5784e+003./(fi-7)+1.5870e+003./(fi-8)-1.5956e+003./(fi-9)+1.6385e+003./(fi-10)-1.6213e+003./(fi-11)+1.6642e+003./(fi-12)-1.6728e+003./(fi-13)+1.7157e+003./(fi-14)-1.7157e+003./(fi-15)+1.7071e+003./(fi-16)-1.7157e+003./(fi-17)+1.7071e+003./(fi-18)-1.6813e+003./(fi-19)+1.6728e+003./(fi-20)-1.6813e+003./(fi-21)+1.6470e+003./(fi-22)-1.6299e+003./(fi-23)+1.5870e+003./(fi-24)-1.5269e+003./(fi-25)+1.4926e+003./(fi-26)-1.5184e+003./(fi-27)+1.4755e+003./(fi-28)-1.5012e+003./(fi-29)+1.5355e+003./(fi-30)-1.5870e+003./(fi-31)+1.5698e+003./(fi-32)-1.6041e+003./(fi-33)+1.5870e+003./(fi-34)-1.6041e+003./(fi-35)+1.5870e+003./(fi-36)-1.5784e+003./(fi-37)+1.5784e+003./(fi-38)-1.5870e+003./(fi-39)+1.5098e+003./(fi-40)-1.5012e+003./(fi-41)+1.5012e+003./(fi-42)-1.4755e+003./(fi-43)+7.4631e+002./(fi-44)+7.4631e+002./ffi)./(8.5783*(1./(fi-2)-1./(fi-1)-1./(fi-3)+1./(fi-4)-1./(fi-5)+1./(fi-6)-1./(fi-7)+1./(fi-8)-1./(fi-9)+1./(fi-10)-1./(fi-11)+1./(fi-12)-1./(fi-13)+1./(fi-14)-1./(fi-15)+1./(fi-16)-1./(fi-17)+1./(fi-18)-1./(fi-19)+1./(fi-20)-1./(fi-21)+1./(fi-22)-1./(fi-23)+1./(fi-24)+1./(fi-25)-1./(fi-26)+1./(fi-27)-1./(fi-28)+1./(fi-29)-1./(fi-30)+1./(fi-31)-1./(fi-32)+1./(fi-33)-1./(fi-34)+1./(fi-35)-1./(fi-36)+1./(fi-37)-1./(fi-38)+1./(fi-39)-1./(fi-40)+1./(fi-41)-1./(fi-42)+1./(fi-43)-1./(fi-44)+1./(fi-45)-1./(fi-46)+1./(fi-47)-1./(fi-48)+1./(fi-49)-1./(fi-50)+0.5./(fi-51)-0.5./ffi)).*(3.2137e+002./(fi-1)-3.2321e+002./(fi-2)+3.1952e+002./(fi-3)-3.1398e+002./(fi-4)+3.1583e+002./(fi-5)-3.1767e+002./(fi-6)+3.2321e+002./(fi-7)-3.2691e+002./(fi-8)+3.3245e+002./(fi-9)-3.3429e+002./(fi-10)+3.4353e+002./(fi-11)-3.4538e+002./(fi-12)+3.4353e+002./(fi-13)-3.4538e+002./(fi-14)+3.3984e+002./(fi-15)-3.3429e+002./(fi-16)+3.2506e+002./(fi-17)-3.2506e+002./(fi-18)+3.1952e+002./(fi-19)-3.1583e+002./(fi-20)+3.1398e+002./(fi-21)-3.0659e+002./(fi-22)+2.8997e+002./(fi-23)-2.7519e+002./(fi-24)+2.7335e+002./(fi-25)-2.8443e+002./(fi-26)+2.8258e+002./(fi-27)-2.8812e+002./(fi-28)+3.0844e+002./(fi-29)-3.1583e+002./(fi-30)+3.1952e+002./(fi-31)-3.2506e+002./(fi-32)+3.2137e+002./(fi-33)-3.2875e+002./(fi-34)+3.3060e+002./(fi-35)-3.3614e+002./(fi-36)+3.4353e+002./(fi-37)-3.4538e+002./(fi-38)+3.4907e+002./(fi-39)-3.4722e+002./(fi-40)+3.4722e+002./(fi-41)-3.4538e+002./(fi-42)+3.4538e+002./(fi-43)-3.4722e+002./(fi-44)+3.4168e+002./(fi-45)-3.4168e+002./(fi-46)+3.4168e+002./(fi-47)-3.3799e+002./(fi-48)+3.4168e+002./(fi-49)-3.3799e+002./(fi-50)+1.6068e+002./(fi-51)-1.6068e+002./ffi); (8-33)

$VS_{BLI}^{c}{}^{W_8}(f_x)$=1./(1.8469*(1./(fi-1)-1./(fi-2)+1./(fi-3)-1./(fi-4)+1./(fi-5)-1./(fi-6)+1./(fi-7)-1./(fi-8)+1./(fi-9)-1./(fi-10)+1./(fi-11)-1./(fi-12)+1./(fi-13)-1./(fi-14)+1./(fi-15)-1./(fi-16)+1./(fi-17)-1./(fi-18)+1./(fi-19)-1./(fi-20)+1./(fi-21)-1./(fi-22)+1./(fi-23)-1./(fi-24)+1./(fi-25)-1./(fi-26)+1./(fi-27)-1./(fi-28)+1./(fi-29)-1./(fi-30)+1./(fi-31)-1./(fi-32)+1./(fi-33)-1./(fi-34)+1./(fi-35)-1./(fi-36)+1./(fi-37)-1./(fi-38)+1./(fi-39)-1./(fi-40)+1./(fi-41)-1./(fi-42)+1./(fi-43)-1./(fi-44)+1./(fi-45)-1./(fi-46)+1./(fi-47)-1./(fi-48)+1./(fi-49)-1./(fi-50)+0.5./(fi-51)-0.5./ffi)).*(3.2137e+002./(fi-1)-3.2321e+002./(fi-2)+3.1952e+002./(fi-3)-3.1398e+002./(fi-4)+3.1583e+002./(fi-5)-3.1767e+002./(fi-6)+3.2321e+002./(fi-7)-3.2691e+002./(fi-8)+3.3245e+002./(fi-9)-3.3429e+002./(fi-10)+3.4353e+002./(fi-11)-3.4538e+002./(fi-12)+3.4353e+002./(fi-13)-3.4538e+002./(fi-14)+3.3984e+002./(fi-15)-3.3429e+002./(fi-16)+3.2506e+002./(fi-17)-3.2506e+002./(fi-18)+3.1952e+002./(fi-19)-3.1583e+002./(fi-20)+3.1398e+002./(fi-21)-3.0659e+002./(fi-22)+2.8997e+002./(fi-23)-2.7519e+002./(fi-24)+2.7335e+002./(fi-25)-2.8443e+002./(fi-26)+2.8258e+002./(fi-27)-2.8812e+002./(fi-28)+3.0844e+002./(fi-29)-3.1583e+002./(fi-30)+3.1952e+002./(fi-31)-3.2506e+002./(fi-32)+3.2137e+002./(fi-33)-3.2875e+002./(fi-34)+3.3060e+002./(fi-35)-3.3614e+002./(fi-36)+3.4353e+002./(fi-37)-3.4538e+002./(fi-38)+3.4907e+002./(fi-39)-3.4722e+002./(fi-40)+3.4722e+002./(fi-41)-3.4538e+002./(fi-42)+3.4538e+002./(fi-43)-3.4722e+002./(fi-44)+3.4168e+002./(fi-45)-3.4168e+002./(fi-46)+3.4168e+002./(fi-47)-3.3799e+002./(fi-48)+3.4168e+002./(fi-49)-3.3799e+002./(fi-50)+1.6068e+002./(fi-51)-1.6068e+002./ffi); (8-34)

$VS_{BLI}^{c}{}^{W_9}(f_x)$=1./(1.8469*(1./(fi-1)-1./(fi-2)+1./(fi-3)-1./(fi-4)+1./(fi-5)-1./(fi-6)+1./(fi-7)-1./(fi-8)+1./(fi-9)-1./(fi-10)+1./(fi-11)-1./(fi-12)+1./(fi-13)-1./(fi-14)+1./(fi-15)-1./(fi-16)+1./(fi-17)-1./(fi-18)+1./(fi-19)-1./(fi-20)+1./(fi-21)-1./(fi-22)+1./(fi-23)-1./(fi-24)+1./(fi-25)-1./(fi-26)+1./(fi-27)-1./(fi-28)+1./(fi-29)-1./(fi-30)+1./(fi-31)-1./(fi-32)+1./(fi-33)-1./(fi-34)+1./(fi-35)-1./(fi-36)+1./(fi-37)-1./(fi-38)+1./(fi-39)-1./(fi-40)+1./(fi-41)-1./(fi-42)+1./(fi-43)-1./(fi-44)+1./(fi-45)-1./(fi-46)+1./(fi-47)-1./(fi-48)+1./(fi-49)-1./(fi-50)+0.5./(fi-51)-0.5./ffi)).*(3.2691e+002./(fi-1)-3.2506e+002./(fi-2)+3.2321e+002./(fi-3)-3.2321e+002./(fi-4)+3.1767e+002./(fi-5)-3.1583e+002./(fi-6)+3.1583e+002./(fi-7)-3.1583e+002./(fi-8)+3.2321e+002./(fi-9)-3.2506e+002./(fi-10)+3.2137e+002./(fi-11)-3.2321e+002./(fi-12)+3.2321e+002./(fi-13)-3.0844e+002./(fi-14)+3.2506e+002./(fi-15)-3.1398e+002./(fi-16)+3.3245e+002./(fi-17)-3.3245e+002./(fi-18)+3.3060e+002./(fi-19)-3.2137e+002./(fi-20)+3.2321e+002./(fi-21)-3.1398e+002./(fi-22)+3.1398e+002./(fi-23)-3.1952e+002./(fi-24)+3.2875e+002./(fi-25)-3.2875e+002./(fi-26)+3.2691e+002./(fi-27)-3.3429e+002./(fi-28)+3.3245e+002./(fi-29)-3.2875e+002./(fi-30)+3.2875e+002./(fi-31)-3.3429e+002./(fi-32)+3.2875e+002./(fi-33)-3.2321e+002./(fi-34)+3.2506e+002./(fi-35)-3.2875e+002./(fi-36)+3.1583e+002./(fi-37)-3.1583e+002./(fi-38)+3.1213e+002./(fi-39)-3.2137e+002./(fi-40)+3.1767e+002./(fi-41)-3.2506e+002./(fi-42)+3.2137e+002./(fi-43)-3.2137e+002./(fi-44)+3.2506e+002./(fi-45)-3.2875e+002./(fi-46)+3.1952e+002./(fi-47)-3.1767e+002./(fi-48)+3.1583e+002./(fi-49)-3.2321e+002./(fi-50)+1.6068e+002./(fi-51)-1.6068e+002./ffi); (8-35)

$VS_{BLI}^{c}{}^{W_{10}}(f_x)$=1./(3.6980*(1./(fi-1)-1./(fi-2)+1./(fi-3)-1./(fi-4)+1./(fi-5)-1./(fi-6)+1./(fi-7)-1./(fi-8)+1./(fi-9)-1./(fi-10)+1./(fi-11)-1./(fi-12)+1./(fi-13)-1./(fi-14)+1./(fi-15)-1./(fi-16)+1./(fi-17)-1./(fi-18)+1./(fi-19)-1./(fi-20)+1./(fi-21)-1./(fi-22)+1./(fi-23)-1./(fi-24)+1./(fi-25)-1./(fi-26)+1./(fi-27)-1./(fi-28)+1./(fi-29)-1./(fi-30)+1./(fi-31)-1./(fi-32)+1./(fi-33)-1./(fi-34)+1./(fi-35)-1./(fi-36)+1./(fi-37)-1./(fi-38)+1./(fi-39)-1./(fi-40)+1./(fi-41)-1./(fi-42)+1./(fi-43)-1./(fi-44)+1./(fi-45)-1./(fi-46)+1./(fi-47)-1./(fi-48)+1./(fi-49)-1./(fi-50)+1./(fi-51)-1./(fi-52)+0.5./(fi-53)-0.5./ffi)).*(6.5085e+002./(fi-1)-6.5455e+002./(fi-2)+6.3976e+002./(fi-3)-6.3236e+002./(fi-4)+6.3976e+002./(fi-5)-6.2866e+002./(fi-6)+6.4345e+002./(fi-7)-6.5085e+002./(fi-8)+6.6934e+002./(fi-9)-6.8043e+002./(fi-10)+6.8043e+002./(fi-11)-6.7304e+002./(fi-12)+6.6564e+002./(fi-13)-6.6194e+002./(fi-14)+6.6194e+002./(fi-15)-6.6934e+002./(fi-16)+6.5085e+002./(fi-17)-6.5085e+002./(fi-18)+6.5085e+002./(fi-19)-6.4715e+002./(fi-20)+6.4715e+002./(fi-21)-6.3236e+002./(fi-22)+6.1387e+002./(fi-23)-6.2496e+002./(fi-24)+6.2127e+002./(fi-25)-6.1017e+002./(fi-26)+5.9908e+002./(fi-27)-6.1757e+002./(fi-28)+6.1017e+002./(fi-29)-6.2496e+002./(fi-30)+6.3976e+002./(fi-31)-6.5085e+002./(fi-32)+6.5825e+002./(fi-33)-6.5085e+002./(fi-34)+6.5825e+002./(fi-35)-6.5455e+002./(fi-36)+6.6194e+002./(fi-37)-6.6934e+002./(fi-38)+6.6194e+002./(fi-39)-6.6564e+002./(fi-40)+6.7304e+002./(fi-41)-6.5825e+002./(fi-42)+6.7304e+002./(fi-43)-6.6934e+002./(fi-44)+6.4715e+002./(fi-45)-6.4715e+002./(fi-46)+6.6194e+002./(fi-47)-6.6194e+002./(fi-48)+6.5085e+002./(fi-49)-6.4345e+002./(fi-50)+6.3236e+002./(fi-51)-6.5455e+002./(fi-52)+3.2173e+002./(fi-53)-3.2173e+002./ffi); (8-36)

$VS_{BLI}^{c}{}^{W_{11}}(f_x)$=1./(8.5283*(1./(fi-1)-1./(fi-2)+1./(fi-3)-1./(fi-4)+1./(fi-5)-1./(fi-6)+1./(fi-7)-1./(fi-8)+1./(fi-9)-1./(fi-10)+1./(fi-11)-1./(fi-12)+1./(fi-13)-1./(fi-14)+1./(fi-15)-1./(fi-16)+1./(fi-17)-1./(fi-18)+1./(fi-19)-1./(fi-20)+1./(fi-21)-1./(fi-22)+1./(fi-23)-1./(fi-24)+1./(fi-25)-1./(fi-26)+1./(fi-27)-1./(fi-28)+1./(fi-29)-1./(fi-30)+1./(fi-31)-1./(fi-32)+1./(fi-33)-1./(fi-34)+1./(fi-35)-1./(fi-36)+1./(fi-37)-1./(fi-38)+1./(fi-39)-1./(fi-40)+1./(fi-41)-1./(fi-42)+1./(fi-43)-1./(fi-44)+1./(fi-45)-1./(fi-46)+1./(fi-47)-1./(fi-48)+0.5./(fi-49)-0.5./ffi)).*(1.5010e+003./(fi-1)-1.4498e+003./(fi-2)+1.4669e+003./(fi-3)-1.4669e+003./(fi-4)+1.5095e+003./(fi-5)-1.4925e+003./(fi-6)+1.4839e+003./(fi-7)-1.4583e+003./(fi-8)+1.5095e+003./(fi-9)-1.5095e+003./(fi-10)+1.5436e+003./(fi-11)-1.5180e+003./(fi-12)+1.4925e+003./(fi-13)-1.5692e+003./(fi-14)+1.5266e+003./(fi-15)-1.5266e+003./(fi-16)+1.5180e+003./(fi-17)-1.5095e+003./(fi-18)+1.5095e+003./(fi-19)-1.4839e+003./(fi-20)+1.4754e+003./(fi-21)-1.5180e+003./(fi-22)+1.5010e+003./(fi-23)-1.4839e+003./(fi-24)+1.4754e+003./(fi-25)-1.5266e+003./(fi-26)+1.4669e+003./(fi-27)-1.4754e+003./(fi-28)+1.3560e+003./(fi-29)-1.2963e+003./(fi-30)+1.4157e+003./(fi-31)-1.3901e+003./(fi-32)+1.3901e+003./(fi-33)-1.4669e+003./(fi-34)+1.3986e+003./(fi-35)-1.4583e+003./(fi-36)+1.4072e+003./(fi-37)-1.4583e+003./(fi-38)+1.4328e+003./(fi-39)-1.4754e+003./(fi-40)+1.5010e+003./(fi-41)-1.4413e+003./(fi-42)+1.4413e+003./(fi-43)-1.4669e+003./(fi-44)+1.4839e+003./(fi-45)-1.4328e+003./(fi-46)+1.5010e+003./(fi-47)-1.5095e+003./(fi-48)+7.4196e+002./(fi-49)-7.4196e+002./ffi); (8-37)

$VS_{BLI}^{c}{}^{W_{12}}(f_x)$=1./(9.0949*(1./(fi-2)-1./(fi-1)-1./(fi-3)+1./(fi-4)-1./(fi-5)+1./(fi-6)-1./(fi-7)+1./(fi-8)-1./(fi-9)+1./(fi-10)-1./(fi-11)+1./(fi-12)-1./(fi-13)+1./(fi-14)-1./(fi-15)+1./(fi-16)-1./(fi-17)+1./(fi-18)-1./(fi-19)+1./(fi-20)-1./(fi-21)+1./(fi-22)-1./(fi-23)+1./(fi-24)-1./(fi-25)+1./(fi-26)-1./(fi-27)+1./(fi-28)-1./(fi-29)+1./(fi-30)-1./(fi-31)+1./(fi-32)-1./(fi-33)+1./(fi-34)-1./(fi-35)+1./(fi-36)-1./(fi-37)+1./(fi-38)-1./(fi-39)+0.5./(fi-40)+0.5./ffi)).*(1.5734e+003./(fi-2)-1.5734e+003./(fi-1)-1.5552e+003./(fi-3)+1.5643e+003./(fi-4)-1.5370e+003./(fi-5)+1.5734e+003./(fi-6)-1.6007e+003./(fi-7)+1.6280e+003./(fi-8)-1.6644e+003./(fi-9)+1.6644e+003./(fi-10)-1.6644e+003./(fi-11)+1.6735e+003./(fi-12)-1.6826e+003./(fi-13)+1.7008e+003./(fi-14)-1.7008e+003./(fi-15)+1.6826e+003./(fi-16)-1.6826e+003./(fi-17)+1.6644e+003./(fi-18)-1.6371e+003./(fi-19)+1.5734e+003./(fi-20)-1.5825e+003./(fi-21)+1.5734e+003./(fi-22)-1.5461e+003./(fi-23)+1.5825e+003./(fi-24)-1.5916e+003./(fi-25)+1.5825e+003./(fi-26)-1.5825e+003./(fi-27)+1.5916e+003./(fi-28)-1.6007e+003./(fi-29)+1.6007e+003./(fi-30)-1.6098e+003./(fi-31)+1.6189e+003./(fi-32)-1.6371e+003./(fi-33)+1.5916e+003./(fi-34)-1.6280e+003./(fi-35)+1.5916e+003./(fi-36)-1.5916e+003./(fi-37)+1.5825e+003./(fi-38)-1.6098e+003./(fi-39)+7.9126e+002./(fi- (8-38)

40)+7.9126e+002./ffi);

$VS_{BLI}^{c\ W_{13}}(f_x)$=(8.7739e+002./(fi-1)-8.6725e+002./(fi-2)+8.7739e+002./(fi-3)-8.7232e+002./(fi-4)+8.6725e+002./(fi-5)-8.5203e+002./(fi-6)+8.7232e+002./(fi-7)-8.6725e+002./(fi-8)+8.7232e+002./(fi-9)-8.7739e+002./(fi-10)+8.9260e+002./(fi-11)-8.8246e+002./(fi-12)+8.8753e+002./(fi-13)-8.7232e+002./(fi-14)+8.6725e+002./(fi-15)-8.7232e+002./(fi-16)+8.5710e+002./(fi-17)-8.9260e+002./(fi-18)+8.8246e+002./(fi-19)-8.9260e+002./(fi-20)+8.8753e+002./(fi-21)-8.7232e+002./(fi-22)+8.5710e+002./(fi-23)-8.4696e+002./(fi-24)+8.6725e+002./(fi-25)-8.7232e+002./(fi-26)+8.5203e+002./(fi-27)-8.6725e+002./(fi-28)+8.5203e+002./(fi-29)-8.4696e+002./(fi-30)+8.5710e+002./(fi-31)-8.5710e+002./(fi-32)+8.5710e+002./(fi-33)-8.5710e+002./(fi-34)+8.5710e+002./(fi-35)-8.6217e+002./(fi-36)+8.3682e+002./(fi-37)-8.3682e+002./(fi-38)+8.2160e+002./(fi-39)-8.5710e+002./(fi-40)+8.6725e+002./(fi-41)-8.7232e+002./(fi-42)+4.4123e+002./(fi-43)-4.4123e+002./ffi)./(5.0716*(1./(fi-1)-1./(fi-2)+1./(fi-3)-1./(fi-4)+1./(fi-5)-1./(fi-6)+1./(fi-7)-1./(fi-8)+1./(fi-9)-1./(fi-10)+1./(fi-11)-1./(fi-12)+1./(fi-13)-1./(fi-14)+1./(fi-15)-1./(fi-16)+1./(fi-17)-1./(fi-18)+1./(fi-19)-1./(fi-20)+1./(fi-21)-1./(fi-22)+1./(fi-23)-1./(fi-24)+1./(fi-25)-1./(fi-26)+1./(fi-27)-1./(fi-28)+1./(fi-29)-1./(fi-30)+1./(fi-31)-1./(fi-32)+1./(fi-33)-1./(fi-34)+1./(fi-35)-1./(fi-36)+1./(fi-37)-1./(fi-38)+1./(fi-39)-1./(fi-40)+1./(fi-41)-1./(fi-42)+0.5./(fi-43)-0.5./ffi)); (8-39)

d. NORMALIZED RATIOS OF THE UPPER TO CORNER VISUAL SPEECH SIGNALS

$VS_{BLI}^{uc\ W_1}(f_x)$=-(1.8723./(ffi - 2) - 1.5303./(ffi - 1) - 2.0099./(ffi - 3) + 3.1104./(ffi - 4) - 3.6607./(ffi - 5) + 3.5033./(ffi - 6) - 3.2042./(ffi - 7) + 3.6980./(ffi - 8) - 3.3840./(ffi - 9) + 3.4655./(ffi - 10) - 3.2862./(ffi - 11) + 2.3344./(ffi - 12) - 2.4629./(ffi - 13) + 2.2489./(ffi - 14) - 1.9460./(ffi - 15) + 1.5584./(ffi - 16) - 1.5972./(ffi - 17) + 1.5057./(ffi - 18) - 1.3373./(ffi - 19) + 1.9460./(ffi - 20) - 1.8570./(ffi - 21) + 9.1235e-001./(ffi - 22) - 8.5243e-001./(ffi - 23) + 7.3553e-001./(ffi - 24) - 9.5933e-001./(ffi - 25) + 5.3411e-001./(ffi - 26) + 5.7398e-001./(ffi - 28) - 4.9999e-001./(ffi - 29) + 8.5562e-001./(ffi - 30) - 1.0230./(ffi - 31) + 1.1011./(ffi - 32) - 1.5972./(ffi - 33) + 2.2850./(ffi - 34) - 2.4226./(ffi - 35) + 2.1079./(ffi - 36) - 1.9735./(ffi - 37) + 2.1553./(ffi - 38) - 2.0649./(ffi - 39) + 2.2866./(ffi - 40) - 2.1514./(ffi - 41) + 2.1985./(ffi - 42) - 2.7393./(ffi - 43) + 2.4674./(ffi - 44) - 2.1922./(ffi - 45) + 2.3767./(ffi - 46) - 2.2423./(ffi - 47) + 2.1514./(ffi - 48) - 1.8391./(ffi - 49) + 1.9591./(ffi - 50) - 1.3403./(ffi - 51) + 1.5156./(ffi - 52) - 5.8353e-001./(ffi - 53) + 5.8353e-001./ffi)./(3.6980*(1./(ffi - 1) - 1./(ffi - 2) + 1./(ffi - 3) - 1./(ffi - 4) + 1./(ffi - 5) - 1./(ffi - 6) + 1./(ffi - 7) - 1./(ffi - 8) + 1./(ffi - 9) - 1./(ffi - 10) + 1./(ffi - 11) - 1./(ffi - 12) + 1./(ffi - 13) - 1./(ffi - 14) + 1./(ffi - 15) - 1./(ffi - 16) + 1./(ffi - 17) - 1./(ffi - 18) + 1./(ffi - 19) - 1./(ffi - 20) + 1./(ffi - 21) - 1./(ffi - 22) + 1./(ffi - 23) - 1./(ffi - 24) + 1./(ffi - 25) - 1./(ffi - 26) + 1./(ffi - 27) - 1./(ffi - 28) + 1./(ffi - 29) - 1./(ffi - 30) + 1./(ffi - 31) - 1./(ffi - 32) + 1./(ffi - 33) - 1./(ffi - 34) + 1./(ffi - 35) - 1./(ffi - 36) + 1./(ffi - 37) - 1./(ffi - 38) + 1./(ffi - 39) - 1./(ffi - 40) + 1./(ffi - 41) - 1./(ffi - 42) + 1./(ffi - 43) - 1./(ffi - 44) + 1./(ffi - 45) - 1./(ffi - 46) + 1./(ffi - 47) - 1./(ffi - 48) + 1./(ffi - 49) - 1./(ffi - 50) + 1./(ffi - 51) - 1./(ffi - 52) + 0.5./(ffi - 53) - 0.5./ffi)); (8-40)

$VS_{BLI}^{uc\ W_2}(f_x)$=-(2.2484./(ffi - 2) - 2.7973./(ffi - 1) - 3.9637./(ffi - 3) + 6.3962./(ffi - 4) - 7.1053./(ffi - 5) + 7.6612./(ffi - 6) - 7.8065./(ffi - 7) + 7.5176./(ffi - 8) - 7.1006./(ffi - 9) + 7.1006./(ffi - 10) - 6.1315./(ffi - 11) + 5.1729./(ffi - 12) - 4.9187./(ffi - 13) + 3.9822./(ffi - 14) - 3.2868./(ffi - 15) + 3.4044./(ffi - 16) - 4.7605./(ffi - 17) + 4.4727./(ffi - 18) - 3.6433./(ffi - 19) + 2.9307./(ffi - 20) - 2.6296./(ffi - 21) + 2.2102./(ffi - 22) + 1.0418e-01./(ffi - 24) - 5.3302e-01./(ffi - 25) + 1.4106./(ffi - 26) - 1.7328./(ffi - 27) + 1.6429./(ffi - 28) - 2.7271./(ffi - 29) + 4.8874./(ffi - 30) - 6.0955./(ffi - 31) + 5.7987./(ffi - 32) - 5.8163./(ffi - 33) + 6.8226./(ffi - 34) - 8.1019./(ffi - 35) + 8.5283./(ffi - 36) - 8.0806./(ffi - 37) + 7.9533./(ffi - 38) - 7.3944./(ffi - 39) + 7.3944./(ffi - 40) - 7.6755./(ffi - 41) + 7.2419./(ffi - 42) - 6.8226./(ffi - 43) + 6.9632./(ffi - 44) - 6.2573./(ffi - 45) + 5.5647./(ffi - 46) - 4.5612./(ffi - 47) + 4.1372./(ffi - 48) - 1.4594./(ffi - 49) + 1.4594./ffi)./(8.5283*(1./(ffi - 1) - 1./(ffi - 2) + 1./(ffi - 3) - 1./(ffi - 4) + 1./(ffi - 5) - 1./(ffi - 6) + 1./(ffi - 7) - 1./(ffi - 8) + 1./(ffi - 9) - 1./(ffi - 10) + 1./(ffi - 11) - 1./(ffi - 12) + 1./(ffi - 13) - 1./(ffi - 14) + 1./(ffi - 15) - 1./(ffi - 16) + 1./(ffi - 17) - 1./(ffi - 18) + 1./(ffi - 19) - 1./(ffi - 20) + 1./(ffi - 21) - 1./(ffi - 22) + 1./(ffi - 23) - 1./(ffi - 24) + 1./(ffi - 25) - 1./(ffi - 26) + 1./(ffi - 27) - 1./(ffi - 28) + 1./(ffi - 29) - 1./(ffi - 30) + 1./(ffi - 31) - 1./(ffi - 32) + 1./(ffi - 33) - 1./(ffi - 34) + 1./(ffi - 35) - 1./(ffi - 36) + 1./(ffi - 37) - 1./(ffi - 38) + 1./(ffi - 39) - 1./(ffi - 40) + 1./(ffi - 41) - 1./(ffi - 42) + 1./(ffi - 43) - 1./(ffi - 44) + 1./(ffi - 45) - 1./(ffi - 46) + 1./(ffi - 47) - 1./(ffi - 48) + 0.5./(ffi - 49) - 0.5./ffi)); (8-41)

$VS_{BLI}^{uc\ W_3}(f_x)$=1./(3.6292 *(1./(ffi - 1) - 1./(ffi - 2) + 1./(ffi - 3) - 1./(ffi - 4) + 1./(ffi - 5) - 1./(ffi - 6) + 1./(ffi - 7) - 1./(ffi - 8) + 1./(ffi - 9) - 1./(ffi - 10) + 1./(ffi - 11) - 1./(ffi - 12) + 1./(ffi - 13) - 1./(ffi - 14) + 1./(ffi - 15) - 1./(ffi - 16) + 1./(ffi - 17) - 1./(ffi - 18) + 1./(ffi - 19) - 1./(ffi - 20) + 1./(ffi - 21) - 1./(ffi - 22) + 1./(ffi - 23) - 1./(ffi - 24) + 1./(ffi - 25) - 1./(ffi - 26) + 1./(ffi - 27) - 1./(ffi - 28) + 1./(ffi - 29) - 1./(ffi - 30) + 1./(ffi - 31) - 1./(ffi - 32) + 1./(ffi - 33) - 1./(ffi - 34) + 1./(ffi - 35) - 1./(ffi - 36) + 1./(ffi - 37) - 1./(ffi - 38) + 1./(ffi - 39) - 1./(ffi - 40) + 1./(ffi - 41) - 1./(ffi - 42) + 1./(ffi - 43) - 1./(ffi - 44) + 1./(ffi - 45) - 1./(ffi - 46) + 0.5./(ffi - 47) - 0.5./ffi)).*(1.0226./(ffi - 1) - 1.4571./(ffi - 2) + 1.9420./(ffi - 3) - 2.9681./(ffi - 4) + 3.1378./(ffi - 5) - 3.5823./(ffi - 6) + 3.5332./(ffi - 7) - 3.6292./(ffi - 8) + 3.5332./(ffi - 9) - 3.0587./(ffi - 10) + 2.5358./(ffi - 11) - 2.1491./(ffi - 12) + 1.9936./(ffi - 13) - 1.6079./(ffi - 14) + 1.6800./(ffi - 15) - 2.0319./(ffi - 16) + 1.6635./(ffi - 17) - 2.2325./(ffi - 18) + 9.9577e-001./(ffi - 19) - 5.1579e-001./(ffi - 20) + 8.0584e-081./(ffi - 21) - 2.9819e-002./(ffi - 22) + 8.0584e-081./(ffi - 23) - 4.8316e-001./(ffi - 24) + 7.4758e-001./(ffi - 25) - 1.1786./(ffi - 26) + 1.0026./(ffi - 27) - 3.1952./(ffi - 28) + 3.3932./(ffi - 29) - 3.4394./(ffi - 30) + 3.2048./(ffi - 31) - 3.3221./(ffi - 32) + 3.2501./(ffi - 33) - 2.8962./(ffi - 34) + 3.0292./(ffi - 35) - 2.7028./(ffi - 36) + 2.4244./(ffi - 37) - 2.3091./(ffi - 38) + 2.3064./(ffi - 39) - 2.3503./(ffi - 40) + 1.8723./(ffi - 41) - 1.6800./(ffi - 42) + 1.7091./(ffi - 43) - 1.7537./(ffi - 44) + 1.7912./(ffi - 45) - 1.2108./(ffi - 46) + 5.3766e-001./(ffi - 47) - 5.3766e-001./ffi); (8-42)

$VS_{BLI}^{uc\ W_4}(f_x)$=(5.9328e-001./(ffi - 1) - 9.2891e-001./(ffi - 2) + 1.2652./(ffi - 3) - 2.2346./(ffi - 4) + 2.7620./(ffi - 5) - 3.6292./(ffi - 6) + 3.4873./(ffi - 7) - 3.3727./(ffi - 8) + 3.3851./(ffi - 9) - 3.2108./(ffi - 10) + 3.1966./(ffi - 11) - 2.7631./(ffi - 12) + 2.5371./(ffi - 13) - 2.0924./(ffi - 14) + 2.0924./(ffi - 15) - 2.2788./(ffi - 16) + 2.1602./(ffi - 17) - 2.1947./(ffi - 18) + 2.1947./(ffi - 19) - 2.8017./(ffi - 20) + 2.8412./(ffi - 21) - 2.0249./(ffi - 22) + 1.0888./(ffi - 23) - 2.5860e-001./(ffi - 24) + 8.0584e-006./(ffi - 25) - 4.4920e-001./(ffi - 26) + 4.7060e-001./(ffi - 27) - 9.7667e-001./(ffi - 28) + 1.4622./(ffi - 29) - 2.0391./(ffi - 30) + 2.0733./(ffi - 31) - 2.0324./(ffi - 32) + 1.8776./(ffi - 33) - 1.7534./(ffi - 34) + 1.5628./(ffi - 35) - 1.5944./(ffi - 36) + 1.2771./(ffi - 37) - 1.2771./(ffi - 38) + 1.3374./(ffi - 39) - 1.4029./(ffi - 40) + 1.3991./(ffi - 41) - 1.3017./(ffi - 42) + 1.3017./(ffi - 43) - 1.0682./(ffi - 44) + 1.0387./(ffi - 45) - 9.9894e-001./(ffi - 46) + 3.6482e-001./(ffi - 47) - 3.6482e-001./ffi)./(3.6292*(1./(ffi - 1) - 1./(ffi - 2) + 1./(ffi - 3) - 1./(ffi - 4) + 1./(ffi - 5) - 1./(ffi - 6) + 1./(ffi - 7) - 1./(ffi - 8) + 1./(ffi - 9) - 1./(ffi - 10) + 1./(ffi - 11) - 1./(ffi - 12) + 1./(ffi - 13) - 1./(ffi - 14) + 1./(ffi - 15) - 1./(ffi - 16) + 1./(ffi - 17) - 1./(ffi - 18) + 1./(ffi - 19) - 1./(ffi - 20) + 1./(ffi - 21) - 1./(ffi - 22) + 1./(ffi - 23) - 1./(ffi - 24) + 1./(ffi - 25) - 1./(ffi - 26) + 1./(ffi - 27) - 1./(ffi - 28) + 1./(ffi - 29) - 1./(ffi - 30) + 1./(ffi - 31) - 1./(ffi - 32) + 1./(ffi - 33) - 1./(ffi - 34) + 1./(ffi - 35) - 1./(ffi - 36) + 1./(ffi - 37) - 1./(ffi - 38) + 1./(ffi - 39) - 1./(ffi - 40) + 1./(ffi - 41) - 1./(ffi - 42) + 1./(ffi - 43) - 1./(ffi - 44) + 1./(ffi - 45) - 1./(ffi - 46) + 0.5./(ffi - 47) - 0.5./ffi)); (8-43)

$VS_{BLI}^{uc\ W_5}(f_x)$=(5.1392e-002./(ffi - 1) - 3.2248e-001./(ffi - 2) + 1.3371e-001./(ffi - 3) - 4.3574e-002./(ffi - 4) + 2.3968e-001./(ffi - 5) - 9.5597e-001./(ffi - 6) + 1.3022./(ffi - 7) - 1.6765./(ffi - 8) + 1.6814./(ffi - 9) - 1.7866./(ffi - 10) + 1.9077./(ffi - 11) - 1.8322./(ffi - 12) + 1.6693./(ffi - 13) - 1.4466./(ffi - 14) + 1.2238./(ffi - 15) - 9.7705e-001./(ffi - 16) + 9.0202e-001./(ffi - 17) - 1.2289./(ffi - 18) + 9.0202e-001./(ffi - 19) - 4.0617e-001./(ffi - 20) + 5.7725e-002./(ffi - 21) + 1.0170e-001./(ffi - 23) - 1.9168e-001./(ffi - 24) + 3.2694e-001./(ffi - 25) - 2.0407e-001./(ffi - 26) + 1.5689e-001./(ffi - 27) - 2.2806e-001./(ffi - 28) + 4.7754e-001./(ffi - 29) - 1.0569./(ffi - 30) + 1.1891./(ffi - 31) - 8.7634e-001./(ffi - 32) + 7.4958e-001./(ffi - 33) - 9.5422e-001./(ffi - 34) + 1.2831./(ffi - 35) - 1.3667./(ffi - 36) + 9.7947e-001./(ffi - 37) - 1.0837./(ffi - 38) + 1.1616./(ffi - 39) - 1.4240./(ffi - 40) + 1.1891./(ffi - 41) - 1.3312./(ffi - 42) + 1.1939./(ffi - 43) - 1.2734./(ffi - 44) + 1.3022./(ffi - 45) - 1.3457./(ffi - 46) + 1.3952./(ffi - 47) - 1.2734./(ffi - 48) + 1.2221./(ffi - 49) - 1.3685./(ffi - 50) + 1.2169./(ffi - 51) - 1.0369./(ffi - 52) + 1.1462./(ffi - 53) - 9.5690e-001./(ffi - 54) + 7.6324e-001./(ffi - 55) - 9.0202e-001./(ffi - 56) + (8-44)

5.1342e-001./(ffi - 57) - 3.7157e-001./(ffi - 58) + 7.8446e-002./(ffi - 59) - 7.8446e-002./ffi)./(1.9077*(1./(ffi - 1) - 1./(ffi - 2) + 1./(ffi - 3) - 1./(ffi - 4) + 1./(ffi - 5) - 1./(ffi - 6) + 1./(ffi - 7) - 1./(ffi - 8) + 1./(ffi - 9) - 1./(ffi - 10) + 1./(ffi - 11) - 1./(ffi - 12) + 1./(ffi - 13) - 1./(ffi - 14) + 1./(ffi - 15) - 1./(ffi - 16) + 1./(ffi - 17) - 1./(ffi - 18) + 1./(ffi - 19) - 1./(ffi - 20) + 1./(ffi - 21) - 1./(ffi - 22) + 1./(ffi - 23) - 1./(ffi - 24) + 1./(ffi - 25) - 1./(ffi - 26) + 1./(ffi - 27) - 1./(ffi - 28) + 1./(ffi - 29) - 1./(ffi - 30) + 1./(ffi - 31) - 1./(ffi - 32) + 1./(ffi - 33) - 1./(ffi - 34) + 1./(ffi - 35) - 1./(ffi - 36) + 1./(ffi - 37) - 1./(ffi - 38) + 1./(ffi - 39) - 1./(ffi - 40) + 1./(ffi - 41) - 1./(ffi - 42) + 1./(ffi - 43) - 1./(ffi - 44) + 1./(ffi - 45) - 1./(ffi - 46) + 1./(ffi - 47) - 1./(ffi - 48) + 1./(ffi - 49) - 1./(ffi - 50) + 1./(ffi - 51) - 1./(ffi - 52) + 1./(ffi - 53) - 1./(ffi - 54) + 1./(ffi - 55) - 1./(ffi - 56) + 1./(ffi - 57) - 1./(ffi - 58) + 0.5./(ffi - 59) - 0.5./ffi));

$VS_{BLI}^{uc\ W_6}(f_x)$=(1.2016./(ffi - 1) - 8.0493e-001./(ffi - 2) + 1.7501./(ffi - 3) - 1.6811./(ffi - 4) + 2.8063./(ffi - 5) - 3.8347./(ffi - 6) + 4.6758./(ffi - 7) - 5.0716./(ffi - 8) + 4.7429./(ffi - 9) - 4.6985./(ffi - 10) + 3.9486./(ffi - 11) - 3.3688./(ffi - 12) + 3.1369./(ffi - 13) - 2.6896./(ffi - 14) + 2.5378./(ffi - 15) - 2.8349./(ffi - 16) + 2.5735./(ffi - 17) - 2.1094./(ffi - 18) + 2.1865./(ffi - 19) - 1.5218./(ffi - 20) + 2.9407e-001./(ffi - 21) - 6.5812e-001./(ffi - 22) + 5.3257e-002./(ffi - 23) - 5.6306e-074./(ffi - 24) + 1.0446e-001./(ffi - 25) - 1.6871./(ffi - 26) + 2.7959./(ffi - 27) - 3.1847./(ffi - 28) + 2.9235./(ffi - 29) - 2.9770./(ffi - 30) + 3.1099./(ffi - 31) - 3.4476./(ffi - 32) + 3.4476./(ffi - 33) - 3.5280./(ffi - 34) + 4.2212./(ffi - 35) - 4.0541./(ffi - 36) + 3.4316./(ffi - 37) - 2.4808./(ffi - 38) + 1.5476./(ffi - 39) - 2.0059./(ffi - 40) + 1.0754./(ffi - 41) - 1.1549./(ffi - 42) + 5.2197e-001./(ffi - 43) - 5.2197e-001./ffi)./(5.0716*(1./(ffi - 1) - 1./(ffi - 2) + 1./(ffi - 3) - 1./(ffi - 4) + 1./(ffi - 5) - 1./(ffi - 6) + 1./(ffi - 7) - 1./(ffi - 8) + 1./(ffi - 9) - 1./(ffi - 10) + 1./(ffi - 11) - 1./(ffi - 12) + 1./(ffi - 13) - 1./(ffi - 14) + 1./(ffi - 15) - 1./(ffi - 16) + 1./(ffi - 17) - 1./(ffi - 18) + 1./(ffi - 19) - 1./(ffi - 20) + 1./(ffi - 21) - 1./(ffi - 22) + 1./(ffi - 23) - 1./(ffi - 24) + 1./(ffi - 25) - 1./(ffi - 26) + 1./(ffi - 27) - 1./(ffi - 28) + 1./(ffi - 29) - 1./(ffi - 30) + 1./(ffi - 31) - 1./(ffi - 32) + 1./(ffi - 33) - 1./(ffi - 34) + 1./(ffi - 35) - 1./(ffi - 36) + 1./(ffi - 37) - 1./(ffi - 38) + 1./(ffi - 39) - 1./(ffi - 40) + 1./(ffi - 41) - 1./(ffi - 42) + 0.5./(ffi - 43) - 0.5./ffi)); (8-45)

$VS_{BLI}^{uc\ W_7}(f_x)$=(3.0027./(ffi - 2) - 3.1858./(ffi - 1) - 2.5261./(ffi - 3) + 2.6661./(ffi - 4) - 4.3121./(ffi - 5) + 6.0748./(ffi - 6) - 6.3134./(ffi - 7) + 7.0556./(ffi - 8) - 7.5098./(ffi - 9) + 7.2546./(ffi - 10) - 7.4653./(ffi - 11) + 6.1412./(ffi - 12) - 5.5105./(ffi - 13) + 5.0563./(ffi - 14) - 3.7540./(ffi - 15) + 3.8364./(ffi - 16) - 3.7540./(ffi - 17) + 3.3129./(ffi - 18) - 4.3547./(ffi - 19) + 4.4419./(ffi - 20) - 4.0889./(ffi - 21) + 4.1664./(ffi - 22) - 3.5209./(ffi - 23) + 4.8029./(ffi - 24) - 1.9775./(ffi - 25) + 1.1164./(ffi - 26) + 1.9048e-005./(ffi - 27) + 9.7363e-001./(ffi - 28) - 1.4463e-001./(ffi - 29) + 1.6048./(ffi - 30) - 4.2397./(ffi - 31) + 4.9937./(ffi - 32) - 6.2876./(ffi - 33) + 6.4924./(ffi - 34) - 7.4019./(ffi - 35) + 8.1820./(ffi - 36) - 8.5783./(ffi - 37) + 8.2952./(ffi - 38) - 6.2108./(ffi - 39) + 3.6238./(ffi - 40) - 3.7168./(ffi - 41) + 3.1214./(ffi - 42) - 3.3966./(ffi - 43) + 1.0073./(ffi - 44) + 1.0073./ffi)./(8.5783 *(1./(ffi - 2) - 1./(ffi - 1) - 1./(ffi - 3) + 1./(ffi - 4) - 1./(ffi - 5) + 1./(ffi - 6) - 1./(ffi - 7) + 1./(ffi - 8) - 1./(ffi - 9) + 1./(ffi - 10) - 1./(ffi - 11) + 1./(ffi - 12) - 1./(ffi - 13) + 1./(ffi - 14) - 1./(ffi - 15) + 1./(ffi - 16) - 1./(ffi - 17) + 1./(ffi - 18) - 1./(ffi - 19) + 1./(ffi - 20) - 1./(ffi - 21) + 1./(ffi - 22) - 1./(ffi - 23) + 1./(ffi - 24) - 1./(ffi - 25) + 1./(ffi - 26) - 1./(ffi - 27) + 1./(ffi - 28) - 1./(ffi - 29) + 1./(ffi - 30) - 1./(ffi - 31) + 1./(ffi - 32) - 1./(ffi - 33) + 1./(ffi - 34) - 1./(ffi - 35) + 1./(ffi - 36) - 1./(ffi - 37) + 1./(ffi - 38) - 1./(ffi - 39) + 1./(ffi - 40) - 1./(ffi - 41) + 1./(ffi - 42) - 1./(ffi - 43) + 0.5./(ffi - 44) + 0.5./ffi)); (8-46)

$VS_{BLI}^{uc\ W_8}(f_x)$=UC8=-1./(1.8469 *(1./(ffi - 1) - 1./(ffi - 2) + 1./(ffi - 3) - 1./(ffi - 4) + 1./(ffi - 5) - 1./(ffi - 6) + 1./(ffi - 7) - 1./(ffi - 8) + 1./(ffi - 9) - 1./(ffi - 10) + 1./(ffi - 11) - 1./(ffi - 12) + 1./(ffi - 13) - 1./(ffi - 14) + 1./(ffi - 15) - 1./(ffi - 16) + 1./(ffi - 17) - 1./(ffi - 18) + 1./(ffi - 19) - 1./(ffi - 20) + 1./(ffi - 21) - 1./(ffi - 22) + 1./(ffi - 23) - 1./(ffi - 24) + 1./(ffi - 25) - 1./(ffi - 26) + 1./(ffi - 27) - 1./(ffi - 28) + 1./(ffi - 29) - 1./(ffi - 30) + 1./(ffi - 31) - 1./(ffi - 32) + 1./(ffi - 33) - 1./(ffi - 34) + 1./(ffi - 35) - 1./(ffi - 36) + 1./(ffi - 37) - 1./(ffi - 38) + 1./(ffi - 39) - 1./(ffi - 40) + 1./(ffi - 41) - 1./(ffi - 42) + 1./(ffi - 43) - 1./(ffi - 44) + 1./(ffi - 45) - 1./(ffi - 46) + 1./(ffi - 47) - 1./(ffi - 48) + 1./(ffi - 49) - 1./(ffi - 50) + 0.5./(ffi - 51) - 0.5./ffi)).*(5.0047e-001./(ffi - 2) - 4.2705e-001./(ffi - 1) - 5.1690e-001./(ffi - 3) + 9.2169e-001./(ffi - 4) - 1.2734./(ffi - 5) + 1.6498./(ffi - 6) - 1.7090./(ffi - 7) + 1.5314./(ffi - 8) - 1.8469./(ffi - 9) + 1.5929./(ffi - 10) - 1.6386./(ffi - 11) + 1.3843./(ffi - 12) - 1.3843./(ffi - 13) + 1.2614./(ffi - 14) - 1.0755./(ffi - 15) + 1.1972./(ffi - 16) - 9.9879e-001./(ffi - 17) + 3.8036e-001./(ffi - 18) + 4.1010e-088./(ffi - 19) + 1.0072e-002./(ffi - 20) - 2.8558e-001./(ffi - 21) + 1.2567./(ffi - 22) - 1.0782./(ffi - 23) + 1.1938./(ffi - 24) - 1.1504./(ffi - 25) + 1.0935./(ffi - 26) - 8.2727e-001./(ffi - 27) + 7.5651e-001./(ffi - 28) - 6.0524e-001./(ffi - 29) + 6.3872e-001./(ffi - 30) - 8.6614e-001./(ffi - 31) + 8.6833e-001./(ffi - 32) - 1.0935./(ffi - 33) + 8.0603e-001./(ffi - 34) - 7.8796e-001./(ffi - 35) + 3.5446e-001./(ffi - 36) - 2.8863e-001./(ffi - 37) + 4.0699e-001./(ffi - 38) - 3.8003e-001./(ffi - 39) + 3.0087e-001./(ffi - 40) - 3.3198e-001./(ffi - 41) + 3.8547e-001./(ffi - 42) - 4.1719e-001./(ffi - 43) + 3.5903e-001./(ffi - 44) - 3.2797e-001./(ffi - 45) + 4.1131e-001./(ffi - 46) - 4.9183e-001./(ffi - 47) + 5.6733e-001./(ffi - 48) - 4.6617e-001./(ffi - 49) + 5.1690e-001./(ffi - 50) - 2.3768e-001./(ffi - 51) + 2.3768e-001./ffi); (8-47)

$VS_{BLI}^{uc\ W_9}(f_x)$=-(8.2112e-092./(ffi - 1) + 1.5491e-001./(ffi - 2) - 3.5035e-001./(ffi - 3) + 1.0549./(ffi - 4) - 1.1762./(ffi - 5) + 1.2733./(ffi - 6) - 1.6007./(ffi - 7) + 2.0745./(ffi - 8) - 2.4238./(ffi - 9) + 2.5725./(ffi - 10) - 2.4000./(ffi - 11) + 2.6484./(ffi - 12) - 2.5495./(ffi - 13) + 2.2332./(ffi - 14) - 1.7898./(ffi - 15) + 1.4591./(ffi - 16) - 1.5333./(ffi - 17) + 1.5333./(ffi - 18) - 1.5333./(ffi - 19) + 1.5668./(ffi - 20) - 1.5668./(ffi - 21) + 1.2406./(ffi - 22) - 9.3021e-001./(ffi - 23) + 2.7325e-001./(ffi - 24) - 2.0616e-001./(ffi - 25) - 8.2112e-092./(ffi - 26) - 4.7213e-001./(ffi - 27) + 1.1839./(ffi - 28) - 1.5393./(ffi - 29) + 2.1519./(ffi - 30) - 2.6443./(ffi - 31) + 3.2470./(ffi - 32) - 3.4295./(ffi - 33) + 3.6980./(ffi - 34) - 3.3403./(ffi - 35) + 3.3836./(ffi - 36) - 3.0314./(ffi - 37) + 2.4238./(ffi - 38) - 2.0559./(ffi - 39) + 1.5794./(ffi - 40) - 9.9116e-001./(ffi - 41) + 1.1107./(ffi - 42) - 6.4227e-001./(ffi - 43) + 5.8203e-001./(ffi - 44) - 8.4114e-001./(ffi - 45) + 6.5972e-001./(ffi - 46) - 6.3691e-001./(ffi - 47) + 6.3691e-001./(ffi - 48) - 7.2156e-001./(ffi - 49) + 5.0592e-001./(ffi - 50) - 3.1226e-001./(ffi - 51) + 2.4460e-001./(ffi - 52) - 7.0498e-002./(ffi - 53) + 7.0498e-002./ffi)./(3.6980* (1./(ffi - 1) - 1./(ffi - 2) + 1./(ffi - 3) - 1./(ffi - 4) + 1./(ffi - 5) - 1./(ffi - 6) + 1./(ffi - 7) - 1./(ffi - 8) + 1./(ffi - 9) - 1./(ffi - 10) + 1./(ffi - 11) - 1./(ffi - 12) + 1./(ffi - 13) - 1./(ffi - 14) + 1./(ffi - 15) - 1./(ffi - 16) + 1./(ffi - 17) - 1./(ffi - 18) + 1./(ffi - 19) - 1./(ffi - 20) + 1./(ffi - 21) - 1./(ffi - 22) + 1./(ffi - 23) - 1./(ffi - 24) + 1./(ffi - 25) - 1./(ffi - 26) + 1./(ffi - 27) - 1./(ffi - 28) + 1./(ffi - 29) - 1./(ffi - 30) + 1./(ffi - 31) - 1./(ffi - 32) + 1./(ffi - 33) - 1./(ffi - 34) + 1./(ffi - 35) - 1./(ffi - 36) + 1./(ffi - 37) - 1./(ffi - 38) + 1./(ffi - 39) - 1./(ffi - 40) + 1./(ffi - 41) - 1./(ffi - 42) + 1./(ffi - 43) - 1./(ffi - 44) + 1./(ffi - 45) - 1./(ffi - 46) + 1./(ffi - 47) - 1./(ffi - 48) + 1./(ffi - 49) - 1./(ffi - 50) + 1./(ffi - 51) - 1./(ffi - 52) + 0.5./(ffi - 53) - 0.5./ffi)); (8-48)

$VS_{BLI}^{uc\ W_{10}}(f_x)$= (4.3036e-001./(ffi - 1) - 5.1055e-001./(ffi - 2) + 6.7970e-001./(ffi - 3) - 1.0129./(ffi - 4) + 1.1871./(ffi - 5) - 1.2168./(ffi - 6) + 1.3039./(ffi - 7) - 1.5449./(ffi - 8) + 1.4880./(ffi - 9) - 1.4244./(ffi - 10) + 1.1163./(ffi - 11) - 1.0566./(ffi - 12) + 7.6498e-001./(ffi - 13) - 7.0722e-001./(ffi - 14) + 7.0484e-001./(ffi - 15) - 6.1214e-001./(ffi - 16) + 4.0422e-001./(ffi - 17) - 3.1141e-001./(ffi - 18) + 2.5482e-001./(ffi - 19) - 2.7966e-001./(ffi - 20) + 4.0518e-075./(ffi - 21) + 2.1121e-001./(ffi - 23) - 6.5182e-001./(ffi - 24) + 8.8959e-001./(ffi - 25) - 7.2905e-001./(ffi - 26) + 8.5280e-001./(ffi - 27) - 6.4338e-001./(ffi - 28) + 4.7786e-001./(ffi - 29) - 4.7071e-001./(ffi - 30) + 1.0102./(ffi - 31) - 1.1931./(ffi - 32) + 1.7916./(ffi - 33) - 1.8469./(ffi - 34) + 1.6437./(ffi - 35) - 1.4960./(ffi - 36) + 1.1163./(ffi - 37) - 1.0566./(ffi - 38) + 8.9601e-001./(ffi - 39) - 7.8039e-001./(ffi - 40) + 4.7630e-001./(ffi - 41) - 4.0149e-001./(ffi - 42) + 3.1414e-001./(ffi - 43) - 2.5908e-001./(ffi - 44) + 2.9360e-001./(ffi - 45) - 1.6116e-001./(ffi - 46) + 3.3775e-001./(ffi - 47) - 1.3873e-001./(ffi - 48) + 1.6116e-001./(ffi - 49) - 1.3873e-001./(ffi - 50) + 2.1518e-001./(ffi - 51) - 2.1518e-001./ffi)./(1.8469*(1./(ffi - 1) - 1./(ffi - 2) + 1./(ffi - 3) - 1./(ffi - 4) + 1./(ffi - 5) - 1./(ffi - 6) + 1./(ffi - 7) - 1./(ffi - 8) + 1./(ffi - 9) - 1./(ffi - 10) + 1./(ffi - 11) - 1./(ffi - 12) + 1./(ffi - 13) - 1./(ffi - 14) + 1./(ffi - 15) - 1./(ffi - 16) + 1./(ffi - 17) - 1./(ffi - 18) + 1./(ffi - 19) - 1./(ffi - 20) + 1./(ffi - 21) - 1./(ffi - 22) + 1./(ffi - 23) - 1./(ffi - 24) + 1./(ffi - 25) - 1./(ffi - 26) + 1./(ffi - 27) - 1./(ffi - 28) + 1./(ffi - 29) - 1./(ffi - 30) + 1./(ffi - 31) - 1./(ffi - 32) + 1./(ffi - 33) - 1./(ffi - 34) + 1./(ffi - 35) - 1./(ffi - 36) + 1./(ffi - 37) - 1./(ffi - 38) + 1./(ffi - 39) - 1./(ffi - 40) + 1./(ffi - 41) - 1./(ffi - 42) + 1./(ffi - 43) - 1./(ffi - 44) + 1./(ffi - 45) - 1./(ffi - 46) + 1./(ffi - 47) - 1./(ffi - 48) + 1./(ffi - 49) - 1./(ffi - 50) + 0.5./(ffi - 51) - 0.5./ffi)); (8-49)

$VS_{BLI}^{uc\ W_{11}}(f_x)$=-(1.9727./(ffi - 2) - 2.0511./(ffi - 1) - 2.3492./(ffi - 3) + 2.3492./(ffi - 4) - 1.9787./(ffi - 5) + 3.9293./(ffi - 6) - 4.7918./(ffi - 7) + 7.1759./(ffi - 8) - 5.5478./(ffi - 9) + 6.3127./(ffi - 10) - 5.1873./(ffi - 11) + 6.4702./(ffi - 12) - 5.4765./(ffi - 13) + 5.1724./(ffi - 14) - 4.3572./(ffi - 15) + 3.8530./(ffi - 16) - 4.6957./(ffi - 17) + 5.8028./(ffi - 18) - 5.5478./(ffi - 19) + 2.7171./(ffi - 20) - 4.4757e-001./(ffi - 21) + 1.3252e-001./(ffi - 22) + 6.4244e-001./(ffi - 24) - 1.7517./(ffi - 25) + 8.2789e-001./(ffi - 26) - 4.7104./(ffi - 27) + 4.6209./(ffi - 28) - 2.2865./(ffi - 29) + 8.5283./(ffi - 30) - 7.9882./(ffi - 31) + 6.3956./(ffi - 32) - 5.8419./(ffi - 33) + 5.4974./(ffi - 34) - 5.4655./(ffi - 35) + 4.2731./(ffi - 36) - 4.5466./(ffi - 37) + 4.2731./(ffi - 38) - 3.4673./(ffi - 39) + 3.3167./(ffi - 40) - 2.3075./(ffi - 41) + 3.1161./(ffi - 42) - 3.9171./(ffi - 43) + 3.1363./(ffi - 44) - 2.4578./(ffi - 45) + 2.6616./(ffi - 46) - 2.3075./(ffi - 47) + 1.2139./(ffi - 48) - 9.6956e-001./(ffi - 49) + 9.6956e-001./ffi)./(8.5283 *(1./(ffi - 1) - 1./(ffi - 2) + 1./(ffi - 3) - 1./(ffi - 4) + 1./(ffi - 5) - 1./(ffi - 6) + 1./(ffi - 7) - 1./(ffi - 8) + 1./(ffi - 9) - 1./(ffi - 10) + 1./(ffi - 11) - 1./(ffi - 12) + 1./(ffi - 13) - 1./(ffi - 14) + 1./(ffi - 15) - 1./(ffi - 16) + 1./(ffi - 17) - 1./(ffi - 18) + 1./(ffi - 19) - 1./(ffi - 20) + 1./(ffi - 21) - 1./(ffi - 22) + 1./(ffi - 23) - 1./(ffi - 24) + 1./(ffi - 25) - 1./(ffi - 26) + 1./(ffi - 27) - 1./(ffi - 28) + 1./(ffi - 29) - 1./(ffi - 30) + 1./(ffi - 31) - 1./(ffi - 32) + 1./(ffi - 33) - 1./(ffi - 34) + 1./(ffi - 35) - 1./(ffi - 36) + 1./(ffi - 37) - 1./(ffi - 38) + 1./(ffi - 39) - 1./(ffi - 40) + 1./(ffi - 41) - 1./(ffi - 42) + 1./(ffi - 43) - 1./(ffi - 44) + 1./(ffi - 45) - 1./(ffi - 46) + 1./(ffi - 47) - 1./(ffi - 48) + (8-50)

0.5./(ffi - 49)- 0.5./ffi));

$VS_{BLI}^{uc\ W_{12}}(f_x)$=-(1.3823./(ffi - 1) - 1.3823./(ffi - 2) + 2.2357./(ffi - 3) - 3.4302./(ffi - 4) + 4.4313./(ffi - 5) - 4.9337./(ffi - 6) + 5.8509./(ffi - 7) - 5.8013./(ffi - 8) + 5.9411./(ffi - 9) - 4.7203./(ffi - 10) + 4.4150./(ffi - 11) - 4.9127./(ffi - 12) + 6.0087./(ffi - 13) - 6.3736./(ffi - 14) + 6.6723./(ffi - 15) - 7.2164./(ffi - 16) + 7.8202./(ffi - 17) - 8.9932./(ffi - 18) + 9.0949./(ffi - 19) - 3.9652./(ffi - 20) + 4.1730./(ffi - 21) - 2.3509./(ffi - 22) + 1.3557./(ffi - 23) - 3.5310./(ffi - 24) + 5.3359./(ffi - 25) - 6.4200./(ffi - 26) + 7.3830./(ffi - 27) - 7.8892./(ffi - 28) + 7.1203./(ffi - 29) - 7.1203./(ffi - 30) + 6.0446./(ffi - 31) - 5.2947./(ffi - 32) + 4.4405./(ffi - 33) - 4.3784./(ffi - 34) + 2.9930./(ffi - 35) - 2.4634./(ffi - 36) + 1.8251./(ffi - 37) - 1.6050./(ffi - 38) + 6.8008e-001./(ffi - 39))./(9.0949 *(1./(ffi - 2) - 1./(ffi - 1) - 1./(ffi - 3) + 1./(ffi - 4) - 1./(ffi - 5) + 1./(ffi - 6) - 1./(ffi - 7) + 1./(ffi - 8) - 1./(ffi - 9) + 1./(ffi - 10) - 1./(ffi - 11) + 1./(ffi - 12) - 1./(ffi - 13) + 1./(ffi - 14) - 1./(ffi - 15) + 1./(ffi - 16) - 1./(ffi - 17) + 1./(ffi - 18) - 1./(ffi - 19) + 1./(ffi - 20) - 1./(ffi - 21) + 1./(ffi - 22) - 1./(ffi - 23) + 1./(ffi - 24) - 1./(ffi - 25) + 1./(ffi - 26) - 1./(ffi - 27) + 1./(ffi - 28) - 1./(ffi - 29) + 1./(ffi - 30) - 1./(ffi - 31) + 1./(ffi - 32) - 1./(ffi - 33) + 1./(ffi - 34) - 1./(ffi - 35) + 1./(ffi - 36) - 1./(ffi - 37) + 1./(ffi - 38) - 1./(ffi - 39) + 0.5./(ffi - 40) + 0.5./ffi)); (8-51)

$VS_{BLI}^{uc\ W_{13}}(f_x)$= (2.1623./(ffi - 1) - 2.4855./(ffi - 2) + 2.9720./(ffi - 3) - 3.8500./(ffi - 4) + 4.3286./(ffi - 5) - 4.5481./(ffi - 6) + 4.8680./(ffi - 7) - 4.5334./(ffi - 8) + 5.0716./(ffi - 9) - 4.9963./(ffi - 10) + 3.9796./(ffi - 11) - 3.7142./(ffi - 12) + 3.6475./(ffi - 13) - 5.0716./(ffi - 14) + 4.3286./(ffi - 15) - 9.9952e-001./(ffi - 16) + 1.2245e-001./(ffi - 17) - 1.9898e-001./(ffi - 18) + 4.9401e-001./(ffi - 19) + 3.2473./(ffi - 21) - 4.2572./(ffi - 22) + 4.2668./(ffi - 23) - 4.4133./(ffi - 24) + 3.9190./(ffi - 25) - 3.4428./(ffi - 26) + 3.7142./(ffi - 27) - 3.3047./(ffi - 28) + 3.5058./(ffi - 29) - 3.3647./(ffi - 30) + 3.4380./(ffi - 31) - 3.8524./(ffi - 32) + 3.4380./(ffi - 33) - 3.0235./(ffi - 34) + 3.0235./(ffi - 35) - 3.1649./(ffi - 36) + 3.0775./(ffi - 37) - 3.0775./(ffi - 38) + 2.6334./(ffi - 39) - 2.1946./(ffi - 40) + 1.8711./(ffi - 41) - 2.0175./(ffi - 42) + 9.5143e-001./(ffi - 43) - 9.5143e-001./ffi)./(5.0716 *(1./(ffi - 1) - 1./(ffi - 2) + 1./(ffi - 3) - 1./(ffi - 4) + 1./(ffi - 5) - 1./(ffi - 6) + 1./(ffi - 7) - 1./(ffi - 8) + 1./(ffi - 9) - 1./(ffi - 10) + 1./(ffi - 11) - 1./(ffi - 12) + 1./(ffi - 13) - 1./(ffi - 14) + 1./(ffi - 15) - 1./(ffi - 16) + 1./(ffi - 17) - 1./(ffi - 18) + 1./(ffi - 19) - 1./(ffi - 20) + 1./(ffi - 21) - 1./(ffi - 22) + 1./(ffi - 23) - 1./(ffi - 24) + 1./(ffi - 25) - 1./(ffi - 26) + 1./(ffi - 27) - 1./(ffi - 28) + 1./(ffi - 29) - 1./(ffi - 30) + 1./(ffi - 31) - 1./(ffi - 32) + 1./(ffi - 33) - 1./(ffi - 34) + 1./(ffi - 35) - 1./(ffi - 36) + 1./(ffi - 37) - 1./(ffi - 38) + 1./(ffi - 39) - 1./(ffi - 40) + 1./(ffi - 41) - 1./(ffi - 42) + 0.5./(ffi - 43) - 0.5./ffi)); (8-52)

e. NORMALIZED RATIOS OF THE LOWER TO CORNER VISUAL SPEECH SIGNALS

$VS_{BLI}^{lc\ W_1}(f_x)$=(0.065661./(fi-1)-0.31478./(fi-2)+0.89607./(fi-3)-2.39081./(fi-4)+2.80601./(fi-5)-3.23181./(fi-6)+3.40961./(fi-7)-3.43141./(fi-8)+3.69801./(fi-9)-3.40961./(fi-10)+3.12431./(fi-11)-2.93511./(fi-12)+2.76431./(fi-13)-2.67311./(fi-14)+2.60831./(fi-15)-2.14031./(fi-16)+2.10481./(fi-17)-1.99001./(fi-18)+2.18321./(fi-19)-1.98431./(fi-20)+2.02641./(fi-21)-1.51641./(fi-22)+1.59571./(fi-23)-2.17911./(fi-24)+2.71791./(fi-25)-3.06021./(fi-26)+2.76861./(fi-27)-1.83031./(fi-28)+1.47741./(fi-29)-1.90781./(fi-30)+2.28041./(fi-31)-2.37571./(fi-32)+2.80601./(fi-33)-3.05511./(fi-34)+2.55691./(fi-35)-2.12641./(fi-36)+2.04531./(fi-37)-1.79821./(fi-38)+1.92101./(fi-39)-1.76301./(fi-40)+1.92621./(fi-41)-1.59841./(fi-42)+1.47741./(fi-43)-1.23111./(fi-44)+1.35211./(fi-45)-0.85819./(fi-46)+0.66604./(fi-47)-0.53895./(fi-48)+0.42264./(fi-49)-0.72109./(fi-50)+0.45735./(fi-51)+0.049057./(fi-53)-0.049057./ffi)./(3.69801.*(1./(fi-1)-1./(fi-2)+1./(fi-3)-1./(fi-4)+1./(fi-5)-1./(fi-6)+1./(fi-7)-1./(fi-8)+1./(fi-9)-1./(fi-10)+1./(fi-11)-1./(fi-12)+1./(fi-13)-1./(fi-14)+1./(fi-15)-1./(fi-16)+1./(fi-17)-1./(fi-18)+1./(fi-19)-1./(fi-20)+1./(fi-21)-1./(fi-22)+1./(fi-23)-1./(fi-24)+1./(fi-25)-1./(fi-26)+1./(fi-27)-1./(fi-28)+1./(fi-29)-1./(fi-30)+1./(fi-31)-1./(fi-32)+1./(fi-33)-1./(fi-34)+1./(fi-35)-1./(fi-36)+1./(fi-37)-1./(fi-38)+1./(fi-39)-1./(fi-40)+1./(fi-41)-1./(fi-42)+1./(fi-43)-1./(fi-44)+1./(fi-45)-1./(fi-46)+1./(fi-47)-1./(fi-48)+1./(fi-49)-1./(fi-50)+1./(fi-51)-1./(fi-52)+0.5./(fi-53)-0.5./ffi)); (8-53)

$VS_{BLI}^{lc\ W_2}(f_x)$=(0.46475./(fi-1)-1.7919./(fi-2)+3.2381./(fi-3)-5.4049./(fi-4)+7.2613./(fi-5)-7.8015./(fi-6)+8.5283./(fi-7)-8.4890./(fi-8)+8.2078./(fi-9)-7.9265./(fi-10)+7.7704./(fi-11)-7.0597./(fi-12)+6.7620./(fi-13)-6.3419./(fi-14)+6.7620./(fi-15)-6.9101./(fi-16)+6.5034./(fi-17)-6.0923./(fi-18)+4.6938./(fi-19)-5.6766./(fi-20)+4.6907./(fi-21)-4.6907./(fi-22)+3.9669./(fi-23)-3.8165./(fi-24)+6.4533./(fi-25)-7.3411./(fi-26)+6.8861./(fi-27)-4.6740./(fi-28)+5.8613./(fi-29)-6.6301./(fi-30)+6.2976./(fi-31)-6.1573./(fi-32)+6.2825./(fi-33)-6.8429./(fi-34)+5.6925./(fi-35)-4.2545./(fi-36)+3.5594./(fi-37)-3.2576./(fi-38)+3.8165./(fi-39)-3.5273./(fi-40)+3.1041./(fi-41)-2.4282./(fi-42)+2.5289./(fi-43)-1.4187./(fi-44)+1.2557./(fi-45)-0.73139./(fi-46)+0.39775./(fi-47)-0.11176./(fi-48))./(8.5283.*(1./(fi-1)-1./(fi-2)+1./(fi-3)-1./(fi-4)+1./(fi-5)-1./(fi-6)+1./(fi-7)-1./(fi-8)+1./(fi-9)-1./(fi-10)+1./(fi-11)-1./(fi-12)+1./(fi-13)-1./(fi-14)+1./(fi-15)-1./(fi-16)+1./(fi-17)-1./(fi-18)+1./(fi-19)-1./(fi-20)+1./(fi-21)-1./(fi-22)+1./(fi-23)-1./(fi-24)+1./(fi-25)-1./(fi-26)+1./(fi-27)-1./(fi-28)+1./(fi-29)-1./(fi-30)+1./(fi-31)-1./(fi-32)+1./(fi-33)-1./(fi-34)+1./(fi-35)-1./(fi-36)+1./(fi-37)-1./(fi-38)+1./(fi-39)-1./(fi-40)+1./(fi-41)-1./(fi-42)+1./(fi-43)-1./(fi-44)+1./(fi-45)-1./(fi-46)+1./(fi-47)-1./(fi-48)+0.5./(fi-49)-0.5./ffi)); (8-54)

$VS_{BLI}^{lc\ W_3}(f_x)$=1./((3.6292).*(1./(fi-1)-1./(fi-2)+1./(fi-3)-1./(fi-4)+1./(fi-5)-1./(fi-6)+1./(fi-7)-1./(fi-8)+1./(fi-9)-1./(fi-10)+1./(fi-11)-1./(fi-12)+1./(fi-13)-1./(fi-14)+1./(fi-15)-1./(fi-16)+1./(fi-17)-1./(fi-18)+1./(fi-19)-1./(fi-20)+1./(fi-21)-1./(fi-22)+1./(fi-23)-1./(fi-24)+1./(fi-25)-1./(fi-26)+1./(fi-27)-1./(fi-28)+1./(fi-29)-1./(fi-30)+1./(fi-31)-1./(fi-32)+1./(fi-33)-1./(fi-34)+1./(fi-35)-1./(fi-36)+1./(fi-37)-1./(fi-38)+1./(fi-39)-1./(fi-40)+1./(fi-41)-1./(fi-42)+1./(fi-43)-1./(fi-44)+1./(fi-45)-1./(fi-46)+0.5./(fi-47)-0.5./ffi)).*(0.32046./(fi-1)-0.38729./(fi-2)+0.66012./(fi-3)-2.0970./(fi-4)+2.0571./(fi-5)-2.4322./(fi-6)+2.4078./(fi-7)-2.7431./(fi-8)+2.1650./(fi-9)-2.1650./(fi-10)+1.8387./(fi-11)-1.7987./(fi-12)+1.7995./(fi-13)-1.9178./(fi-14)+1.9200./(fi-15)-1.9200./(fi-16)+2.6723./(fi-17)-3.0848./(fi-18)+2.8563./(fi-19)-2.3796./(fi-20)+1.3242./(fi-21)-0.39250./(fi-22)+0.35677./(fi-23)-0.0.9259./(fi-24)+0.55493./(fi-25)-1.1156./(fi-26)+2.0132./(fi-27)-3.3665./(fi-28)+3.6292./(fi-29)-3.6000./(fi-30)+3.2800./(fi-31)-3.1200./(fi-32)+2.9268./(fi-33)-2.5245./(fi-34)+2.3339./(fi-35)-1.6795./(fi-36)+1.5591./(fi-37)-1.5154./(fi-38)+1.3177./(fi-39)-1.1437./(fi-40)+0.87040./(fi-41)-0.40000./(fi-42)+0.36948./(fi-43)-0.14217./(fi-44)+0.092413./(fi-45)-8.0584e-081./(fi-46)+0.16828./(fi-47)-0.16828./ffi); (8-55)

$VS_{BLI}^{lc\ W_4}(f_x)$=0.41071./(fi-1)-1.0763./(fi-2)+1.0859./(fi-3)-2.8413./(fi-4)+3.1160./(fi-5)-2.8620./(fi-6)+2.8103./(fi-7)-3.0120./(fi-8)+2.8214./(fi-9)-2.3695./(fi-10)+2.2849./(fi-11)-2.3318./(fi-12)+2.3743./(fi-13)-2.2811./(fi-14)+2.1926./(fi-15)-2.0157./(fi-16)+2.1955./(fi-17)-2.5116./(fi-18)+2.6914./(fi-19)-2.9032./(fi-20)+3.4229./(fi-21)-3.6292./(fi-22)+2.5012./(fi-23)-0.32098./(fi-24)+0.79432./(fi-25)-1.1148./(fi-26)+1.1794./(fi-27)-2.4627./(fi-28)+2.7686./(fi-29)-3.1928./(fi-30)+3.1543./(fi-31)-2.7721./(fi-32)+2.4637./(fi-33)-1.7465./(fi-34)+1.3849./(fi-35)-1.3364./(fi-36)+1.2041./(fi-37)-1.0233./(fi-38)+0.64764./(fi-39)-0.51837./(fi-40)+0.54083./(fi-41)-0.44841./(fi-42)+0.44841./(fi-43)-0.43160./(fi-44)+0.11473./(fi-45)+0.15993./(fi-47)-0.15993./ffi)./(3.6292*(1./(fi-1)-1./(fi-2)+1./(fi-3)-1./(fi-4)+1./(fi-5)-1./(fi-6)+1./(fi-7)-1./(fi-8)+1./(fi-9)-1./(fi-10)+1./(fi-11)-1./(fi-12)+1./(fi-13)-1./(fi-14)+1./(fi-15)-1./(fi-16)+1./(fi-17)-1./(fi-18)+1./(fi-19)-1./(fi-20)+1./(fi-21)-1./(fi-22)+1./(fi-23)-1./(fi-24)+1./(fi-25)-1./(fi-26)+1./(fi-27)-1./(fi-28)+1./(fi-29)-1./(fi-30)+1./(fi-31)-1./(fi-32)+1./(fi-33)-1./(fi-34)+1./(fi-35)-1./(fi-36)+1./(fi-37)-1./(fi-38)+1./(fi-39)-1./(fi-40)+1./(fi-41)-1./(fi-42)+1./(fi-43)-1./(fi-44)+1./(fi-45)-1./(fi-46)+0.5./(fi-47)-0.5./ffi)); (8-56)

$VS_{BLI}^{lc\ W_5}(f_x)$=-(2.0111e-001./(fi-2)-0.096474./(fi-1)-1.8317e-001./(fi-3)+3.2533e-001./(fi-4)-6.8071e-001./(fi-5)+1.3239./(fi-6)-1.4030./(fi-7)+1.3445./(fi-8)-1.5105./(fi-9)+1.8134./(fi-10)-1.7488./(fi-11)+1.9077./(fi-12)-1.7545./(fi-13)+1.4808./(fi-14)-1.2462./(fi-15)+1.2502./(fi-16)-1.4083./(fi-17)+1.1927./(fi-18)-1.0998./(fi-19)+6.4896e-001./(fi-20)-5.6332e-001./(fi-21)+6.6650e-001./(fi-22)-6.0223e-001./(fi-23)+1.0571./(fi-24)-8.7782e-001./(fi-25)+7.2279e-001./(fi-26)-7.2431e-001./(fi-27)+1.0104./(fi-28)-1.1275./(fi-29)+1.4669./(fi-30)-1.3499./(fi-31)+1.0377./(fi-32)-1.0759./(fi-33)+1.1025./(fi-34)-1.2364./(fi-35)+9.3668e-001./(fi-36)-8.1432e-001./(fi-37)+7.1158e-001./(fi-38)-8.7561e-001./(fi-39)+0.89634./(fi-40)-0.89634./(fi-41)+1.0034./(fi-42)-1.0437./(fi-43)+1.0018./(fi-44)-9.3940e-001./(fi-45)+0.85511./(fi-46)-7.5259e-001./(fi-47)+8.7606e-001./(fi-48)-8.9725e-001./(fi-49)+8.3343e-001./(fi-50)-0.54419./(fi-51)+3.4367e-001./(fi-52)-0.27649./(fi-53)+0.096474./(fi-54)-0.21926./(fi-55)+0.12309./(fi-56)+0.061188./(fi-58)-0.056957./(fi-59)+0.056957./ffi)./(1.9077*(1./(fi-1)-1./(fi-2)+1./(fi-3)-1./(fi-4)+1./(fi-5)-1./(fi-6)+1./(fi-7)-1./(fi-8)+1./(fi-9)-1./(fi-10)+1./(fi-11)-1./(fi-12)+1./(fi-13)-1./(fi-14)+1./(fi-15)-1./(fi-16)+1./(fi-17)-1./(fi-18)+1./(fi-19)-1./(fi-20)+1./(fi-21)-1./(fi-22)+1./(fi- (8-57)

23)-1./(fi-24)+1./(fi-25)-1./(fi-26)+1./(fi-27)-1./(fi-28)+1./(fi-29)-1./(fi-30)+1./(fi-31)-1./(fi-32)+1./(fi-33)-1./(fi-34)+1./(fi-35)-1./(fi-36)+1./(fi-37)-1./(fi-38)+1./(fi-39)-1./(fi-40)+1./(fi-41)-1./(fi-42)+1./(fi-43)-1./(fi-44)+1./(fi-45)-1./(fi-46)+1./(fi-47)-1./(fi-48)+1./(fi-49)-1./(fi-50)+1./(fi-51)-1./(fi-52)+1./(fi-53)-1./(fi-54)+1./(fi-55)-1./(fi-56)+1./(fi-57)-1./(fi-58)+0.5./(fi-59)-0.5./ffi));

$VS_{BLI}^{lc\ W_1}(f_x)$=(0.12800./(fi-1)+0.50991./(fi-3)-0.24172./(fi-4)+1.8941./(fi-5)-3.8516./(fi-6)+5.0716./(fi-7)-4.2354./(fi-8)+4.5926./(fi-9)-4.5385./(fi-10)+3.8043./(fi-11)-3.3243./(fi-12)+2.0146./(fi-13)-1.3822./(fi-14)+1.2003./(fi-15)-0.94176./(fi-16)+0.98041./(fi-17)-1.3677./(fi-18)+2.0813./(fi-19)-3.1579./(fi-20)+2.7694./(fi-21)-2.3351./(fi-22)+1.3883./(fi-23)-1.2932./(fi-24)+1.9421./(fi-25)-3.4713./(fi-26)+2.7404./(fi-27)-2.7500./(fi-28)+0.87373./(fi-29)-0.31074./(fi-30)+0.32292./(fi-31)-0.95095./(fi-32)+1.7186./(fi-33)-1.8897./(fi-34)+2.0620./(fi-35)-2.4592./(fi-36)+2.2888./(fi-37)-1.8906./(fi-38)+1.0156./(fi-39)-0.93438./(fi-40)+0.63243./(fi-41)-0.39797./(fi-42)+0.10241./(fi-43)-0.10241./ffi)./(5.0716*(1./(fi-1)-1./(fi-2)+1./(fi-3)-1./(fi-4)+1./(fi-5)-1./(fi-6)+1./(fi-7)-1./(fi-8)+1./(fi-9)-1./(fi-10)+1./(fi-11)-1./(fi-12)+1./(fi-13)-1./(fi-14)+1./(fi-15)-1./(fi-16)+1./(fi-17)-1./(fi-18)+1./(fi-19)-1./(fi-20)+1./(fi-21)-1./(fi-22)+1./(fi-23)-1./(fi-24)+1./(fi-25)-1./(fi-26)+1./(fi-27)-1./(fi-28)+1./(fi-29)-1./(fi-30)+1./(fi-31)-1./(fi-32)+1./(fi-33)-1./(fi-34)+1./(fi-35)-1./(fi-36)+1./(fi-37)-1./(fi-38)+1./(fi-39)-1./(fi-40)+1./(fi-41)-1./(fi-42)+0.5./(fi-43)-0.5./ffi)); (8-58)

$VS_{BLI}^{lc\ W_7}(f_x)$=(0.58945./(fi-2)-0.55113./(fi-1)-1.6717./(fi-3)+2.3601./(fi-4)-3.3434./(fi-5)+5.5094./(fi-6)-5.6545./(fi-7)+6.5344./(fi-8)-6.4217./(fi-9)+7.2164./(fi-10)-8.4128./(fi-11)+8.5783./(fi-12)-8.4604./(fi-13)+7.1570./(fi-14)-6.6085./(fi-15)+6.5304./(fi-16)-5.1458./(fi-17)+4.3253./(fi-18)-4.4256./(fi-19)+4.5222./(fi-20)-4.6121./(fi-21)+5.0086./(fi-22)-6.1755./(fi-23)+6.7320./(fi-24)-6.7420./(fi-25)+4.7065./(fi-26)-1.6965./(fi-27)+1.1019./(fi-28)-1.8807./(fi-29)+3.3551./(fi-30)-5.1507./(fi-31)+5.9641./(fi-32)-5.7236./(fi-33)+5.3484./(fi-34)-4.9414./(fi-35)+5.5460./(fi-36)-4.6608./(fi-37)+4.0646./(fi-38)-2.3833./(fi-39)+1.1647./(fi-40)+0.41793./(fi-42)-0.88931./(fi-43)+0.14652./(fi-44)+0.14652./ffi)./(8.5783*(1./(fi-2)-1./(fi-1)-1./(fi-3)+1./(fi-4)-1./(fi-5)+1./(fi-6)-1./(fi-7)+1./(fi-8)-1./(fi-9)+1./(fi-10)-1./(fi-11)+1./(fi-12)-1./(fi-13)+1./(fi-14)-1./(fi-15)+1./(fi-16)-1./(fi-17)+1./(fi-18)-1./(fi-19)+1./(fi-20)-1./(fi-21)+1./(fi-22)-1./(fi-23)+1./(fi-24)-1./(fi-25)+1./(fi-26)-1./(fi-27)+1./(fi-28)-1./(fi-29)+1./(fi-30)-1./(fi-31)+1./(fi-32)-1./(fi-33)+1./(fi-34)-1./(fi-35)+1./(fi-36)-1./(fi-37)+1./(fi-38)-1./(fi-39)+1./(fi-40)-1./(fi-41)+1./(fi-42)-1./(fi-43)+0.5./(fi-44)+0.5./ffi)); (8-59)

$VS_{\sqsubset LI}^{lc\ W_8}(f_x)$=(0.18691./(fi-1)-0.37594./(fi-2)+0.39803./(fi-3)-1.41261./(fi-4)+1.84221./(fi-5)-1.75801./(fi-6)+1.70031./(fi-7)-1.64271./(fi-8)+1.35621./(fi-9)-1.44081./(fi-10)+1.38411./(fi-11)-1.24351./(fi-12)+1.24351./(fi-13)-1.41121./(fi-14)+1.21661./(fi-15)-1.55681./(fi-16)+1.49561./(fi-17)-8.9279e-001./(fi-18)+5.3155e-001./(fi-19)-7.0383e-001./(fi-20)+1.13081./(fi-21)-1.73091./(fi-22)+1.84691./(fi-23)-1.81141./(fi-24)+1.66241./(fi-25)-1.44081./(fi-26)+1.41291./(fi-27)-7.5958e-001./(fi-28)+7.8319e-001./(fi-29)-7.2039e-001./(fi-30)+4.4332e-001./(fi-31)-7.0508e-001./(fi-32)+6.0956e-001./(fi-33)-7.3622e-001./(fi-34)+6.5616e-001./(fi-35)-6.0956e-001./(fi-36)+5.4669e-001./(fi-37)-3.7364e-001./(fi-38)+3.6101e-001./(fi-39)-2.5032e-001./(fi-40)+1.7913e-001./(fi-41)-2.6385e-001./(fi-42)+2.5032e-001./(fi-43)-1.3694e-001./(fi-44)+1.5176e-001./(fi-45)+1.0090e-001./(fi-47)-7.0879e-003./(fi-48)+1.4291e-001./(fi-49)-2.2894e-001./(fi-50)+2.3854e-001./(fi-51)-2.3854e-001./ffi)./(1.84691*(1./(fi-1)-1./(fi-2)+1./(fi-3)-1./(fi-4)+1./(fi-5)-1./(fi-6)+1./(fi-7)-1./(fi-8)+1./(fi-9)-1./(fi-10)+1./(fi-11)-1./(fi-12)+1./(fi-13)-1./(fi-14)+1./(fi-15)-1./(fi-16)+1./(fi-17)-1./(fi-18)+1./(fi-19)-1./(fi-20)+1./(fi-21)-1./(fi-22)+1./(fi-23)-1./(fi-24)+1./(fi-25)-1./(fi-26)+1./(fi-27)-1./(fi-28)+1./(fi-29)-1./(fi-30)+1./(fi-31)-1./(fi-32)+1./(fi-33)-1./(fi-34)+1./(fi-35)-1./(fi-36)+1./(fi-37)-1./(fi-38)+1./(fi-39)-1./(fi-40)+1./(fi-41)-1./(fi-42)+1./(fi-43)-1./(fi-44)+1./(fi-45)-1./(fi-46)+1./(fi-47)-1./(fi-48)+1./(fi-49)-1./(fi-50)+0.5./(fi-51)-0.5./ffi)); (8-60)

$VS_{BLI}^{lc\ W_9}(f_x)$=(0.024010./(fi-1)-0.060367./(fi-2)+0.34476./(fi-3)-8.9810e-001./(fi-4)+1.1477./(fi-5)-1.1797./(fi-6)+1.3010./(fi-7)-1.9667./(fi-8)+2.2679./(fi-9)-2.4524./(fi-10)+2.4524./(fi-11)-2.3497./(fi-12)+2.1259./(fi-13)-1.8033./(fi-14)+1.8033./(fi-15)-1.3824./(fi-16)+1.4204./(fi-17)-1.3596./(fi-18)+1.2989./(fi-19)-1.4526./(fi-20)+1.6968./(fi-21)-2.0853./(fi-22)+2.2110./(fi-23)-1.5910./(fi-24)+1.2440./(fi-25)-1.0197./(fi-26)+1.1828./(fi-27)-1.6606./(fi-28)+2.1205./(fi-29)-2.5394./(fi-30)+2.9388./(fi-31)-3.2417./(fi-32)+3.3979./(fi-33)-3.6059./(fi-34)+3.6980./(fi-35)-3.4409./(fi-36)+3.1762./(fi-37)-2.8583./(fi-38)+2.2808./(fi-39)-1.6510./(fi-40)+1.3516./(fi-41)-1.3569./(fi-42)+1.0581./(fi-43)-9.6921e-001./(fi-44)+9.6415e-001./(fi-45)-9.0310e-001./(fi-46)+6.0941e-001./(fi-47)-3.7064e-001./(fi-48)+3.2756e-001./(fi-49)-3.7990e-001./(fi-50)+2.7324e-001./(fi-51)+0.036429./(fi-53)-0.036429./ffi)./(3.6980*(1./(fi-1)-1./(fi-2)+1./(fi-3)-1./(fi-4)+1./(fi-5)-1./(fi-6)+1./(fi-7)-1./(fi-8)+1./(fi-9)-1./(fi-10)+1./(fi-11)-1./(fi-12)+1./(fi-13)-1./(fi-14)+1./(fi-15)-1./(fi-16)+1./(fi-17)-1./(fi-18)+1./(fi-19)-1./(fi-20)+1./(fi-21)-1./(fi-22)+1./(fi-23)-1./(fi-24)+1./(fi-25)-1./(fi-26)+1./(fi-27)-1./(fi-28)+1./(fi-29)-1./(fi-30)+1./(fi-31)-1./(fi-32)+1./(fi-33)-1./(fi-34)+1./(fi-35)-1./(fi-36)+1./(fi-37)-1./(fi-38)+1./(fi-39)-1./(fi-40)+1./(fi-41)-1./(fi-42)+1./(fi-43)-1./(fi-44)+1./(fi-45)-1./(fi-46)+1./(fi-47)-1./(fi-48)+1./(fi-49)-1./(fi-50)+1./(fi-51)-1./(fi-52)+0.5./(fi-53)-0.5./ffi)); (8-61)

$VS_{BLI}^{lc\ W_{10}}(f_x)$=(0.021930./(fi-1)-0.044012./(fi-2)+0.22383./(fi-3)-0.57756./(fi-4)+0.82377./(fi-5)-1.1423./(fi-6)+1.5588./(fi-7)-1.4390./(fi-8)+1.6953./(fi-9)-1.6002./(fi-10)+1.4531./(fi-11)-1.2933./(fi-12)+1.1403./(fi-13)-1.0512./(fi-14)+0.96987./(fi-15)-1.0287./(fi-16)+0.98425./(fi-17)-7.6383e-001./(fi-18)+6.3494e-001./(fi-19)-6.3472e-001./(fi-20)+8.0575e-001./(fi-21)-9.2627e-001./(fi-22)+1.4567./(fi-23)-1.8469./(fi-24)+1.7458./(fi-25)-1.1994./(fi-26)+9.2830e-001./(fi-27)-8.1941e-001./(fi-28)+7.5041e-001./(fi-29)-7.4815e-001./(fi-30)+1.2703./(fi-31)-1.4618./(fi-32)+1.5083./(fi-33)-1.2711./(fi-34)+1.1774./(fi-35)-1.0089./(fi-36)+1.1055./(fi-37)-1.1550./(fi-38)+1.0812./(fi-39)-8.9436e-001./(fi-40)+9.2875e-001./(fi-41)-7.7464e-001./(fi-42)+5.6718e-001./(fi-43)-5.5045e-001./(fi-44)+4.9635e-001./(fi-45)-3.5656e-001./(fi-46)+1.8181e-001./(fi-47)-1.4132e-001./(fi-48)+7.0662e-003./(fi-49)+0.010965./(fi-51)-0.010965./ffi)./(1.8469*(1./(fi-1)-1./(fi-2)+1./(fi-3)-1./(fi-4)+1./(fi-5)-1./(fi-6)+1./(fi-7)-1./(fi-8)+1./(fi-9)-1./(fi-10)+1./(fi-11)-1./(fi-12)+1./(fi-13)-1./(fi-14)+1./(fi-15)-1./(fi-16)+1./(fi-17)-1./(fi-18)+1./(fi-19)-1./(fi-20)+1./(fi-21)-1./(fi-22)+1./(fi-23)-1./(fi-24)+1./(fi-25)-1./(fi-26)+1./(fi-27)-1./(fi-28)+1./(fi-29)-1./(fi-30)+1./(fi-31)-1./(fi-32)+1./(fi-33)-1./(fi-34)+1./(fi-35)-1./(fi-36)+1./(fi-37)-1./(fi-38)+1./(fi-39)-1./(fi-40)+1./(fi-41)-1./(fi-42)+1./(fi-43)-1./(fi-44)+1./(fi-45)-1./(fi-46)+1./(fi-47)-1./(fi-48)+1./(fi-49)-1./(fi-50)+0.5./(fi-51)-0.5./ffi)); (8-62)

$VS_{BLI}^{lc\ W_{11}}(f_x)$=1./(8.5283.*(1./(fi-1)-1./(fi-2)+1./(fi-3)-1./(fi-4)+1./(fi-5)-1./(fi-6)+1./(fi-7)-1./(fi-8)+1./(fi-9)-1./(fi-10)+1./(fi-11)-1./(fi-12)+1./(fi-13)-1./(fi-14)+1./(fi-15)-1./(fi-16)+1./(fi-17)-1./(fi-18)+1./(fi-19)-1./(fi-20)+1./(fi-21)-1./(fi-22)+1./(fi-23)-1./(fi-24)+1./(fi-25)-1./(fi-26)+1./(fi-27)-1./(fi-28)+1./(fi-29)-1./(fi-30)+1./(fi-31)-1./(fi-32)+1./(fi-33)-1./(fi-34)+1./(fi-35)-1./(fi-36)+1./(fi-37)-1./(fi-38)+1./(fi-39)-1./(fi-40)+1./(fi-41)-1./(fi-42)+1./(fi-43)-1./(fi-44)+1./(fi-45)-1./(fi-46)+1./(fi-47)-1./(fi-48)+0.5./(fi-49)-0.5./ffi)).*(1.8937e-005./(fi-1)-1.9732e-001./(fi-2)+5.1460e-001./(fi-3)-1.0915./(fi-4)+9.2607e-001./(fi-5)-2.7488./(fi-6)+3.5333./(fi-7)-3.4757./(fi-8)+3.4487./(fi-9)-3.3085./(fi-10)+3.1576./(fi-11)-3.2353./(fi-12)+3.4575./(fi-13)-2.8127./(fi-14)+3.4401./(fi-15)-3.8558./(fi-16)+3.3747./(fi-17)-3.0283./(fi-18)+4.4297./(fi-19)-4.6737./(fi-20)+1.8893./(fi-21)-2.1205./(fi-22)+1.4094./(fi-23)-8.2459e-001./(fi-24)+1.0290./(fi-25)-1.3614./(fi-26)+3.3990./(fi-27)-3.6099./(fi-28)+8.5283./(fi-29)-7.4106./(fi-30)+4.6205./(fi-31)-5.0367./(fi-32)+4.5802./(fi-33)-3.9759./(fi-34)+4.3417./(fi-35)-3.1856./(fi-36)+2.7531./(fi-37)-2.0251./(fi-38)+2.3826./(fi-39)-2.0327./(fi-40)+1.8322./(fi-41)-1.8706./(fi-42)+1.2835./(fi-43)-1.2357./(fi-44)+1.1097./(fi-45)-2.2349./(fi-46)+0.70471./(fi-47)-0.92607./(fi-48)+0.12717./(fi-49)-0.12717./ffi); (8-63)

$VS_{BLI}^{lc\ W_{12}}(f_x)$=(0.7292./(fi-2)-0.4492./(fi-1)-1.24831./(fi-3)+1.69131./(fi-4)-4.07271./(fi-5)+4.92921./(fi-6)-6.44061./(fi-7)+6.81881./(fi-8)-5.98031./(fi-9)+6.50971./(fi-10)-7.03911./(fi-11)+6.63201./(fi-12)-5.70561./(fi-13)+5.17781./(fi-14)-5.17781./(fi-15)+5.70561./(fi-16)-5.96751./(fi-17)+7.83321./(fi-18)-9.09491./(fi-19)+7.16921./(fi-20)-5.62441./(fi-21)+1.84921./(fi-22)-3.36451./(fi-23)+5.34601./(fi-24)-5.48121./(fi-25)+7.01631./(fi-26)-7.01631./(fi-27)+5.48121./(fi-28)-4.78921./(fi-29)+3.96351./(fi-30)-2.73671./(fi-31)+2.88431./(fi-32)-2.09811./(fi-33)+2.71321./(fi-34)-2.21841./(fi-35)+2.43641./(fi-36)-1.88281./(fi-37)+1.44861./(fi-38)+0.02831./(fi-40)+0.02831./ffi)./(9.09491.*(1./(fi-2)-1./(fi-1)-1./(fi-3)+1./(fi-4)-1./(fi-5)+1./(fi-6)-1./(fi-7)+1./(fi-8)-1./(fi-9)+1./(fi-10)-1./(fi-11)+1./(fi-12)-1./(fi-13)+1./(fi-14)-1./(fi-15)+1./(fi-16)-1./(fi-17)+1./(fi-18)-1./(fi-19)+1./(fi-20)-1./(fi-21)+1./(fi-22)-1./(fi-23)+1./(fi-24)-1./(fi-25)+1./(fi-26)-1./(fi-27)+1./(fi-28)-1./(fi-29)+1./(fi-30)-1./(fi-31)+1./(fi-32)-1./(fi-33)+1./(fi-34)-1./(fi-35)+1./(fi-36)-1./(fi-37)+1./(fi-38)-1./(fi-39)+0.5./(fi-40)+0.5./ffi)); (8-64)

$VS_{BLI}^{lc\ W_{13}}(f_x)$=-(1.0850./(fi-2)-2.4081./(fi-3)+3.0165./(fi-4)-3.7427./(fi-5)+4.5080./(fi-6)-4.6679./(fi-7)+5.0716./(fi-8)-4.7780./(fi-9)+4.5973./(fi-10)-4.2826./(fi-11)+4.5275./(fi-12)-4.2420./(fi-13)+4.1175./(fi-14)-4.2964./(fi-15)+2.5761./(fi-16)-1.0752./(fi-17)+1.1624./(fi-18)-0.93607./(fi-19)+2.2383./(fi-20)-4.0256./(fi-21)+4.1175./(fi-22)-3.3162./(fi-23)+3.1062./(fi-24)-2.9676./(fi-25)+2.5761./(fi-26)-2.5918./(fi-27)+2.0816./(fi-28)-2.1410./(fi-29)+2.1990./(fi-30)-1.9716./(fi-31)+1.4113./(fi-32)-1.7475./(fi-33)+1.5234./(fi-34)-1.0752./(fi-35)+1.4700./(fi-36)-1.7434./(fi-37)+1.5139./(fi-38)-1.4480./(fi-39)+0.73901./(fi-40)-0.42055./(fi-41)+0.15401./(fi-42)-0.032712./(fi-43)+0.032712./ffi)./(5.0716.*(1./(fi-1)-1./(fi-2)+1./(fi-3)-1./(fi-4)+1./(fi-5)-1./(fi-6)+1./(fi-7)-1./(fi-8)+1./(fi-9)-1./(fi-10)+1./(fi-11)-1./(fi-12)+1./(fi-13)-1./(fi-14)+1./(fi-15)-1./(fi-16)+1./(fi-17)-1./(fi-18)+1./(fi-19)-1./(fi-20)+1./(fi-21)-1./(fi-22)+1./(fi-23)-1./(fi-24)+1./(fi-25)-1./(fi-26)+1./(fi-27)-1./(fi-28)+1./(fi-29)-1./(fi-30)+1./(fi-31)-1./(fi-32)+1./(fi-33)-1./(fi-34)+1./(fi-35)-1./(fi-36)+1./(fi-37)-1./(fi-38)+1./(fi-39)-1./(fi-40)+1./(fi-41)-1./(fi-42)+0.5./(fi-43)-0.5./ffi)); (8-65)

Bibliography

Adjoudani, A. & Benoît, C., 1996. *On the integration of auditory and visual parameters in an HMM-based ASR.* Berlin, Germany, Springer.

Alothmany, N. et al., 2010. *Classification of Visemes Using Visual Cues.* Zadar, s.n.

Ananthakrishnan, G. & Engwall, O., 2008. *Important Regions in the Articulator Trajectory.* Strasbourg, s.n.

Bargmann, R., Blanz, V. & Seidel, H. P., 2008. *A Nonlinear Viseme Model for Triphone-Based Speech Synthesis.* s.l., s.n.

Bartels, R. H., Batty, J. C. & Barsky, B., 1998. An introduction to Splines for Use in Computer Graphics and Geometric modeling. In: *Hermite and Cubic Spline Interpolation.* San Francisco, California, USA: Morgan Kaufmann, pp. 9-17.

Baum, L. E., Petrie, T., Soules, G. & Norman, W., 1970. A Maximization Technique Occurring in the Statistical Analysis of Probabilistic Functions of Markov Chains. *The Annals of Mathematical Statistics,* 41(1970), pp. 164-171.

Benguerel, A. P. & Pichora-Fuller, M. K., 1982. Coarticulation effects in lipreading. *Journal of Speech and Hearing Research,* Volume 25, pp. 600-607.

Berrut, J. P. & Trefethen, L. N., 2004. Barycentric Lagrange. *SIAM REVIEW,* p. 501–517.

Binnie, C. A., Jackson, P. L. & Montgomery, A. A., 1976. Visual intelligibility of consonants: A lipreading screening test with implications for aural rehabilitation. *Journal of Speech and Hearing Disorders,* 41(9), p. 530.

Binnie, C. A., Montgomery, A. A. & Jackson, P. L., 1974. Auditory and visual contributions to the perception of consonants. *Speech, Language and Hearing Research,* Volume 17, pp. 619-630.

Birkholz, P., Kröger, B. J. & Neuschaefer-Rube, C., 2011. Model-Based Reproduction of Articulatory Trajectories for Consonant-Vowel Sequences. *IEEE Transactions on Audio, Speech & Language Processing,* 19(5), pp. 1422-1433.

Blake, A. & Isard, M., 1998. *Active Contours: The Application of Techniques from Graphics, Vision, Control Theory and Statistics to Visual Tracking of Shapes in Motion.* s.l.:Springer-Verlag GmbH.

Bowyer, T., 2006-2013. *A free online Talking Dictionary of English Pronunciation.* USA, Patent No. 20040162719.

Bregler, C., Covell, M. & Slaney, M., 1997. *Video rewrite: Driving visual speech with audio.* Los Angeles, CA, ACM, pp. 353-360.

Bregler, C. & Konig, Y., 1994. *Eigenlips for robust speech recognition.* Adelaide, Australia, IEEE.

Cambridge, U., 2003. *CAMBRIDGE ADVANCED LEARNER'S DICTIONARY.* Cambridge, UK: Cambridge University Press.

Cao, Y., Faloutsos, P., Kohler, E. & Pighin, F., 2004. *Real-Time Speech Motion Synthesis from Recorded Motions.* New York, s.n.

Cappelletta, L. & Harte, N., 2012. *Phoneme-to-Viseme Mapping for Visual Speech Recognition.* Vilamoura, s.n.

Carnicer, J. M., 2008. *Interpolation and reconstruction of curves and surfaces,* s.l.: The University of Zaragoza.

Chadwick, J. E., Haumann, D. R. & Parent, R. E., 1989. *Layered Construction for Deformable Animated Characters.* Boston, Massachusetts, USA, ACM SIGGRAPH.

Chan, M. T., Zhang, Y. & Huang, T. S., 1998. *Real-time lip tracking and bimodal continuous speech recognition.* Redondo Beach, CA, s.n.

Chen, D. T. & Zeltzer, D., 1992. *Pump it up: Computer Animation of a Biomechanically Based Model of Muscle using the Finite Element Method.* Chicago, Illinois, USA, ACM SIGGRAPH.

Chen, T., 2001. Audiovisual speech processing. *IEEE Signal Processing Magazine,* Volume 18, pp. 9-21.

Chen, T. & Rao, R. R., 1998. *Audio-visual integration in multimodal communication.* Utah, USA, IEEE.

Chiou, G. & Hwang, J. N., 1997. Lipreading from color video. *IEEE Transactions on Image Processing,* 6(8), pp. 1192-1195.

Choe, B., Lee, H. & Ko, H. S., 2001. Performance-driven muscle-based facial animation. *Journal of Visualization and Computer Animation,* Volume 12, p. 67–79.

Cohen, M. & Massaro, D., 1993. Modelling coarticulation in synthetic visual speech. In: N. M. Thalman & D. Thalman, eds. *Models and Techniques in Computer Animation.* Tokyo: Springer-Verlag, pp. 139-156.

Conrey, B. L. & Pisoni, D. B., 2003. *Audiovisual Asynchrony Detection for Speech and Nonspeech Signals.* Jorioz, France, s.n.

Cootes, T. F., Edwards, G. J. & Taylor, C. J., 1998. *Active appearance models.* Freiburg, Germany, s.n.

Cootes, T. F., Taylor, C. J., Cooper, D. H. & Graham, J., 1995. Active shape models - their training and application. *Computer Vision and Image Understanding,* 61(1), p. 38–59.

Cosatto, E. & Graf, H., 1998. *Sample-based synthesis of photorealistic talking heads.* s.l., s.n.

Cox, S., Harvey, R. & Newman, J., 2008. *The challenge of multispeaker lip-reading.* Queensland, Australia, International Conference on Auditory-Visual Speech Processing.

Crochiere, R. E. & Rabiner, L. R., 1983. Basic principles of sampling and sampling rate conversion. In: A. V. Oppenheim, ed. *MULTIRATE DIGITAL SIGNAL PROCESSING.* New Jersey: Prentice-Hall, pp. 13-42.

Czup, L., 2000. *Lip representation by image ellipse.* Beijing, China, s.n.

De Martino, J. M., Magalhaes, L. P. & Violaro, F., 2006. Facial animation based on context-dependent visemes. *Journal of Computers and Graphics,* 30(6), pp. 971-980.

de Vega, M., Álvarez, C. & Carreiras, M., 1992. Estudio estadístico de la ortografía castellana: La frecuencia silábica. *Cognitiva,* 4(1), p. 75–114.

Deligne, S., Potamianos, G. & Neti, C., 2002. *Audio-visual speech enhancement with AVCDCN (audio-visual codebook dependent cepstral normalization).* Denver, s.n.

Deng, Z. et al., 2006. Expressive Facial Animation Synthesis By Learning Speech Coarticulation And Expression Spaces. *IEEE Transaction on Visualization and Computer Graphics,* 12(6), pp. 1-12.

Dudley, H., Riesz, R. R. & Watkins, S. A., 1939. A Synthetic Speaker. Volume 227, pp. 739-764.

Dwight, H. B., 1957. *Tables of Integrals and Other Mathematical Data.* New York: The Macmillan Company.

Edwards, G. J., Taylor, C. J. & Cootes, T. F., 1998. Interpreting Face Images Using Active Appearance Models. *International Conference on Automatic Face and Gesture Recognition,* p. 300–305.

Eisert, P., Chaudhuri, S. & Girod, B., 1997. *Speech driven synthesis of talking head sequences.* Erlangen, s.n., pp. 51-56.

Ekman, P. & Friesen, W., 1978. *Facial Action Coding System: A Technique for the Measurement of Facial Movement.* Palo Alto, Consulting Psychologists Press.

Englebienne, G., 2008. *Animating faces from speech.* s.l.:The University of Manchester.

Erber, N. P., Sachs, R. M. & DeFilippo, C. L., 1979 . *Optical synthesis of articulatory images for lip-reading evaluation and instruction.* NY, s.n.

Essa, I. & Pentland, A., 1997. Coding, analysis, interpretation, and recognition of facial expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 19(7), p. 757–763.

Eveno, N., Caplier, A. & Coulon, P. Y., 2002. *Key Points Based Segmentation of Lips.* Lausanne, s.n.

Everitt, B. S. & Skrondal, A., 2010. In: *The Cambridge Dictionary of Statistics.* Cambridge: Cambridge University Press, p. 239.

Everitt, B. S. & Skrondal, A., 2010. In: *The Cambridge Dictionary of Statistics.* 4 ed. Cambridge: Cambridge University Press, p. 397.

Ezzat, T. & Poggio, T., 1998. *MikeTalk: A talking facial display based on morphing visemes.* s.l., s.n.

Ezzat, T. & Poggio, T., 1999. Visual speech synthesis by morphing visemes. *International journal of Computer Vision,* 38(1), p. 45–37.

Faruquie, T. A. et al., 2001. *Audio Driven Facial Animation for Audio-Visual Reality.* Tokyo, Japan, IEEE.

Fisher, C. G., 1968. Confusions among visually perceived consonants. *Journal of Speech and Hearing Research,* 11(4), pp. 796-804.

Fourman, M., 2002. *informatics,* Edinburgh: Routledge.

Garofolo, J. S. et al., 1993. *TIMIT Acoustic-Phonetic Continuous Speech Corpus,* Philadelphia: Linguistic Data Consortium.

Girin, L., Allard, A. & Schwartz, J. L., 2001. *Speech Signals Separation: A New Approach Exploiting the Coherence of Audio and Visual Speech.* Cannes, France, s.n.

Goldschen, A. J., 1993. *Continuous automatic speech recognition by lipreading,* s.l.: Ph.D. dissertation, George Washington University, Washington DC.

Goldschen, A. J., Garcia, O. N. & Petajan, E., 1994. *Continuous optical automatic speech recognition by lipreading.* Pacific Grove, IEEE Computer Society Press.

Govokhina, O., BaillyG., Breton, G. & Bagshaw, P., 2006. *A new trainable trajectory formation system for facial animation.* Athens, s.n.

Graf, H. P., Cosatto, E. & Potamianos, G. G., 1997. *Robust recognition of faces and facial features with a multi-modal system.* Orlando, FL, s.n.

Guenter, B. et al., 1998. *Making faces.* Orlando, Florida, ACM SIGGRAPH.

Gurbuz, S., Tufekci, Z., Patterson, E. & Gowdy, J. N., 2001. *Application of affine-invariant Fourier descriptors to lipreading for audio-visual speech recognition.* Salt Lake City, s.n.

Gutierrez-Osuna, R. et al., 2005. Speech-driven facial animation with realistic dynamics. *IEEE Transactions on Multimedia*, 7(1), p. 33–42.

Hazen, T. J., Saenko, K., La, C. H. & Glass, J. R., 2004. *A segmentbased audio-visual speech recognizer: data collection, development, and initial experiments.* PA, USA, ACM New York.

Heider, F. & Heider, G., 1940. An experimental investigation of lip. *Psychological Monographs,* 52(1), pp. 124-153.

Hill, D. R., Pearce, A. & Wyvill, B., 1988. Animating Speech: An Automated Approach using Speech Synthesised by Rules. *The Visual Computer*, 3(5), p. 277–289.

Hofer, G., Yamagishi, J. & Shimodaira, H., 2008. *Speech-driven lip motion generation with a trajectory HMM.* Brisbane, Australia, s.n.

Huffman, D. A., 1952. *A Method for the Construction of Minimum-Redundancy Codes.* s.l., I.R.E.

Inc., T. M., 2009. *Version 7.8.0.347 (R2009a).* Natick, Massachusetts, United States of America, Patent No. 7,802,268.

International Phonetic Association, C. ©. 2., 2005. *IPA Chart.* [Online] Available at: http://www.langsci.ucl.ac.uk/ipa/ipachart.html

Ip, H. H. S. & Chan, C. S., 1996. Script-Based Facial Gesture and Speech Animation Using a NURBS Based Face Model. *Computers and Graphics,* 20(6), p. 881–891.

ISO/IEC, 1.-2., 2001. *Information technology - Coding of Audio-Visual,* Geneva, Switzerland: ISO.

Jiang, D. M., Ravyse, I., Sahli, H. & Zhang, Y. N., 2008. *Accurate visual speech synthesis based on diviseme unit selection and concatenation.* Queensland, s.n.

Jiang, D., Ravyse, I., Sahli, H. & Verhelst, W., 2008. Speech Driven Realistic Mouth Animation Based on Multi-modal Unit Selection. *Journal of Multi-Modal User Interfaces (Springer),* Volume 2, pp. 157-169.

Jiang, J., Alwan, A., Auer, E. T. & Bernstein, L. E., 2001. *Predicting visual consonant perception from physical measures.* Aalborg, Denmark, Aalborg.

Kalberer, G. A., Mueller, P. & Van Gool, L., 2003 . visual speech, a trajectory in viseme space. *International Journal of Imaging Systems and Technology,* pp. 74-84.

Kalra, P., Mangili, A., Magnenat-Thalmann, N. & Thalmann, D., 1991. *SMILE: A Multilayered Facial Animation System.* Tokyo, Japan, IFIP WG 5.10.

Kass, M., Witkjn, A. & Terzopoulos, D., 1987. Snakes: Active contour models. *International Journal of Computer Vision*, pp. 321-331.

Kass, M., Witkjn, A. & Terzopoulos, D., 1987. Snakes: Active contour models. *International Journal of Computer Vision*, pp. 321-331.

Kim, J. & Davis, C., 2003. *Testing the cuing hypothesis for the AV speech detection.* St Jorioz, France, s.n.

King, S. A., 2001. *A Facial Model and Animation Techniques for Animated Speech.* Ohio: The Ohio State University.

Kobbelt, L. et al., 2000. *Geometric Modeling Based on Polygonal Meshes.* Interlaken, Switzerland, Blackwell Publishers.

Koch, R. M., Gross, M. H. & Bosshard, A. A., 1998. *Emotion Editing using Finite Elements.* Lisbon, Portugal, Eurographics.

Koch, R. M. et al., 1996. *Simulating facial surgery using finite element methods.* New Orleans, LA, Addison-Wesley Reading.

Kshirsagar, S., Escher, M., Sannier, G. & Magnenat-Thalmann, N., 1999. *Multimodal animation system based on the MPEG-4 standard.* Ottawa, Canada, World Scientific.

Kshirsagar, S., Molet, T. & Thalmann, N. M., 2001. *Principal Components of Expressive Speech Animation.* Washington, DC, IEEE Computer Society, p. 38.

Lagrange, J. L., 1877. *Lecons ´elementaires sur les mathématiques, donnees `a l'Ecole Normale en 1795.* Paris: Oeuvres VII, Gauthier–Villars.

Lavagetto, F., 1995. Converting speech into lip movements: A multimedia telephone for hard of hearing people. *IEEE Transactions on Rehabilitation Engineering,* 3(1), pp. 1-14.

Lavagetto, F., Pockaj, R. & Costa, M., 2000. Smooth surface interpolation and texture adaptation for MPEG-4 compliant calibration of 3D head models. *Image and Vision Computing,* 18(4), pp. 345-353.

Lazalde, O. M., Maddock, S. l. & Meredith, M., 2007. A Constraint-Based Approach to Visual Speech for a Mexican-Spanish Talking Head. *International Journal of Computer Games Technology,* Volume 2008, pp. 17-24.

Leawo, S., 2012. *Leawo Video Converter,* Shenzhen: Leawo Software Co. Ltd..

Lee, Y., Terzopoulos, D. & Waters, K., 1993. *Constructing Physics-based Facial Models of Individuals.* Toronto, Ontario, Canada, Graphics Interface.

Lee, Y., Terzopoulos, D. & Waters, K., 1995. *Realistic Modeling for Facial Animations.* Los Angeles, CA, USA, ACM SIGGRAPH.

Lewis, J. P. & Parke, F. I., 1987. *Automated Lip-Synch and Speech Synthesis for Character Animation.* Toronto, Ontario, Proceeding of Human Factors in Computing Systems and Graphics Interface '87.

Lewis, T. W. & Powers, D. M. W., 2003. Audio-visual speech recognition using red exclusion and nerual networks. *Journal of Research and Practice in Information Technology,* 35(1), pp. 41-64.

Ling, Z. H. & Wang, R. H., 2006. *HMM-based unit selection using frame sized speech segments.* Pittsburgh, USA, Kluwer Academic Publishers.

Lin, I. C., Hung, C. S., Yang, T. J. & Ouhyoung, M., 1999. *A speech driven talking head system based on a single face image.* s.l., s.n.

Lucero, J. C., 2002. Identifying a differential equation for lip motion. *Medical Engineering & Physics,* Volume 24, p. 521–528.

Lucey, P., Martin, T. & Sridharan, S., 2004. *Confusability of phonemes grouped according to their viseme classes in noisy environments.* Sydney, Australia, Macquarie University.

Luettin., J. & Thacker, N. A., 1997. Speechreading using probabilistic models. *Computer Vision and Image Understanding,* 65(2), p. 163–178.

Luettin, J., 1997. *Visual Speech and Speaker Recognition.* Sheffield: The University of Sheffield.

Lyons, M. J., Chan, C.-H. & Tetsutani, N., 2004. *MouthType: Text Entry by Hand and Mouth.* Vienna, Austria, s.n.

Magnenat-Thalman, N. & Thalman, D., 1987. The Direction of Synthetic Actors in the Film 'Rendezvous a Montreal. *IEEE journal of Computer Graphics and Applications,* 7(12), pp. 9-19.

Ma, J. et al., 2006. Accurate visible speech synthesis based on concatenating variable length motion capture data. *IEEE Transaction on Visual Computer Graph,* 12(2), p. 1–11.

Mandal, M. & Asif, A., 2007. Part III Discrete-time signals and systems. In: *Continuous and Discrete Time Signals and Systems.* Cambridge: Cambridge University Press, pp. 410-412.

Markov, A., 1913. *An example of statistical investigation in the text of eugene onyegin, illustrating coupling of tests in chains.* s.l., s.n.

Mase, K. & Pentland, A., 1991. Automatic lipreading by optical-flow analysis. *Systems and Computers in Japan,* Volume 22, pp. 67-76.

Massaro, D. W. et al., 1999. *Picture My Voice: Audio to Visual Speech Synthesis using Artificial Neural Networks.* Santa Cruz, CA, s.n.

MathWorks, I., 2008. *MATLAB and Statistics Toolbox 7.7.* Natick, Massachusetts, United States.

Matthews, I. & Baker, S., 2004. Active Appearance Models Revisited. *International Journal of Computer Vision,* 60(2), pp. 135-164.

Mattheyses, W., Latacz, L. & Verhelst, W., 2011. *Automatic viseme clustering for audio visual speech synthesis.* Florence, s.n.

McGurk, H. & MacDonald, J., 1976. Hearing lips and seeing voices. *Nature,* 264(5588), p. 746–748.

Meijering, E., 2002. *A Chronology of Interpolation: From Ancient Astronomy to Modern Signal and Image Processing.* s.l., s.n.

Melenchón, J., Martínez, E., De La Torre, F. & Montero, J. A., 2009. Emphatic Visual Speech Synthesis. *IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING,* 17(3), pp. 459-468.

Mertins, A., 1999. Discrete Signal Representations. In: *Signal Analysis: Wavelets, Filter Banks, Time-Frequency Transforms and applications.* Chichester: John Wiley & Sons, pp. 47-49.

Miller, S., n.d. *The Method of Least Squares,* s.l.: Mathematics Department Brown University Providence.

Montgomery, A. A., 1980. Development of a model for generating synthetic animated lip shapes. *Journal of the Acoustical Society of America,* 68(S1), p. S58.

Morishima, S., Aizawa, K. & Harashima, H., 1989. *An Intelligent Facial Image Coding Driven By Speech And Phoneme.* Glasgow, s.n.

Mortenson, M. E., 1997. B-spline curves. In: M. Spencer, ed. *Geometric Modeling.* s.l.:John Wiley & Sons, Inc, pp. 113-142.

Nahas, M., Huitric, H. & Saintourens, M., 1988. Animation of a B-Spline Figure. *The Visual Computer,* 3(5), p. 272–276.

Nakamura, S., Ito, H. & Shikano, K., 2000. *Stream weight optimization of speech and lip image sequence for audiovisual.* Beijing, s.n.

Nefian, A. R. et al., 2002. Dynamic Bayesian Networks for Audio-Visual Speech Recognition. *EURASIP Journal on Applied Signal Processsing* , 2002(11), pp. 1274-1288.

Neti, C. et al., 2000. *Audio-Visual Speech Recognition,* Baltimore: Center for Language and Speech Processing, The Johns Hopkins University.

Niswar, A., Niswar, E. P., Nguyen, H. T. & Huang, Z., 2009. *Real-time 3D talking head from a synthetic viseme dataset.* Tokyo, ACM, pp. 29-33.

Oetken, G., 1979. Section 8.1. In: *Programs for Digital Signal Processing.* New York: IEEE Press.

Oppenheim, A. V. & Schafer, R. W., 1999. Sampling Of Continuous-time signals. In: T. Robbins, ed. *Discrete-Time Signal Processing.* 3rd ed. New Jersey: Prentice Hall, pp. 150-153.

Otani, K. & Hasegawa, T., 1995. The image input microphone—A new nonacoustic speech communication system by media conversion. *IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS,* p. 42–48.

Oxford, U., 1993. *The New Shorter oxford English Dictionary.* Fourth ed. Oxford: Clarendon Press, Oxford.

Pandzic, I. S. & Forchheimer, R., 2003. *MPEG-4 Facial Animation: The Standard, Implementation and Applications.* s.l.:John Wiley & Sons.

Parke, F. I., 1974. *A Parametric Model for Human Faces,* Utah: The University of Utah.

Parke, F. I., 1982. Parameterized models for facial animation. *IEEE Computer Graphics and Applications,* 2(9), pp. 61-68.

Parke, F. I. & Waters, K., 1996. In: *Computer Facial Animation.* s.l.:A K Peters, p. 188.

Park, F. I., 1972. *Computer generated animation of faces.* New York, s.n.

Pearce, A., Wyvill, B., Wyvill, G. & Hill, D., 1986. *Speech and Expression: A Computer Solution to Face Animation.* Toronto, In Proceedings of the Graphics Interface 86.

Pearce, A., Wyvill, B., Wyvill, G. & Hill, D., 1986. *Speech and Expression: A Computer Solution to Face Animation.* s.l., s.n.

Petajan, E. D., 1984. *Automatic lipreading to enhance speech recognition.* Atlanta, s.n.

Pighin, F. et al., 1998. *Synthesizing realistic facial expressions from photographs.* Orlando, FL, USA, ACM SIGGRAPH.

Platt, S. M. & Badler, N. I., 1981. *Animating Facial Expressions.* s.l., ACM SIGGRAPH.

Potamianos, G. & Graf, H. P., 1998. *Discriminative training of HMM stream exponents for audio-visual speech recognition.* Seattle, WA, s.n.

Potamianos, G., Graf, H. P. & Cosatto, E., 1998. *An image transform approach for HMM based automatic lipreading.* Chicago, USA, IEEE Computer Society.

Potamianos, G., Neti, C., Luettin, J. & Matthews, I., 2004. *Audio-Visual Automatic Speech Recognition: An Overview.* s.l., MIT Press.

Press, T. C. U., 2013. *Cambrige Dictionaries Online.* [Online] Available at: http://dictionary.cambridge.org/dictionary/british/ [Accessed March 2012].

Rabiner, L. & Juang, B. H., 2008. Historical Perspective of the Field of ASR/NLU. In: J. Benesty, M. M. Sondhi & Y. Huang, eds. *Springer Handbook of Speech Processing.* s.l.:Springer, pp. 521-538.

Rabiner, L. & Jung, B. H., 1993. The speech-production process. In: *Fundumentals of speech recognition.* New Jersey: Prentice Hall, p. 14.

Rabiner, L. R., 1989. *A tutorial on hidden Marov models and selected applications in speech recognition.* s.l., IEEE, pp. 257-286.

Ramage, M. D., 2011. *Disproving Visemes As The Basic Visual Unit Of Speech,* Perth: The Curtin University.

Rao, R. & Mersereau, R., 1995. *On Merging Hidden Markov Models with Deformable Templates.* Washington, DC, s.n.

Revéret, L. & Benoît, C., 1998. *A new 3D lip model for analysis and synthesis of lip motion in speech production.* Terrigal, Australia, Auditory-Visual Speech Processing.

Rogozan, A., 1999. Discriminative learning of visual data for audiovisual speech recognition. *International Journal on Artificial Intelligence Tools,* 8(1), p. 43–52.

Rogozan, A. & Deléglise, P. P., 1998. Adaptive fusion of acoustic and visual sources for automatic speech recognition. *Speech Communication,* 26(1-2), p. 149–161.

Rosenblatt, M., 1956. Remarks on some nonparametric estimates of a density function. *The Annals of Mathematical Statistics,* 27(3), pp. 832-837.

Runge, C., 1901. Uber empirische Funktionen und die Interpolation zwischen aquidistanten Ordinaten. *Zeitschrift fur Mathematik und Physik,* p. 224–243.

Saenko, K., Darrell, T. & Glass, J. R., 2004. *Articulatory features for robust visual speech recognition.* PA, USA, ACM.

Saitoh, T. & Konishi, R., 2005. *Lip reading based on sampled active contour model.* Tampa, USA, Springer Berlin Heidelberg.

Saitoh, T. & Konishi, R., 2006. Word recognition based on two dimensional lip motion trajectory. *IEEE International symposium on intelligent signal processing and communication*, pp. 287-290.

Salzer, H. E., 1972. Lagrangian interpolation at the Chebyshev points xn,ν ,= cos(νπ/n), ν = 0(1)n; some unnoted advantages. *Journal of Computer*, p. 156–159.

Scheepers, F., Parent, R. E., Carlson, W. E. & May, S. F., 1997. *Anatomy-Based Modeling of the Human Musculature.* Los Angeles, CA, USA, ACM SIGGRAPH.

Schoenberg, I. J., 1946. Contributions to the problem of approximation of equidistant data by analytic functions. *Quarterly of Applied Mathematics,* Volume 4, p. 45–99.

Seels, B. & Glasgow, Z., 1990. *Exercises in instructional design.* Columbus, Ohio: Charles Merrill.

Seguier, R. & Cladel, N., 2003. *A Multiobjectives Genetic Snakes: Application on Audio-Visual Speech Recognition.* s.l., s.n.

Silva, W. & Oney, L. i., 2009. *GS1 Bar Code Verification for linear symbols.* s.l.:Global Standards.

Silveira, L. G., Facon, J. & Borges, D. L., 2003. *Visual Speech Recognition: a solution from feature extraction to words classification.* Sao Carlos, Brazil, s.n.

Stillittano, S., Girondel, V. & Caplier, A., 2009. *Inner And Outer Lip Contour Tracking Using Cubic Curve Parametric Models.* Cairo, s.n.

Sumby, W. H. & Pollack, I., 1954. Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America,* 26(2), p. 212–215.

Sumby, W. H. & Pollack, I., 1954. Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society,* 26(2), pp. 212-215.

Su, Q. & Silsbee, P. L., 1996. *Robust audiovisual integration using semicontinuous hidden Markov models.* Philadelphia, PA, ISCA.

Technologies, V., n.d. *MPEG-4 Face and Body Animation (MPEG-4 FBA) An overview*, Linköping, Sweden: Visage Technologies AB.

Teissier, P., Robert-Ribes, J. & Schwartz, J. L., 1999. Comparing models for audiovisual fusion in a noisy-vowel recognition task. *IEEE Transactions on Speech and Audio Processing*, 7(6), p. 629–642.

Terzopoulos, D. & Waters, K., 1990. Physically-based facial modeling, analysis and animation. *Journal of Visualization and Computer Animation*, Volume 1, p. 73–80.

Trefethen, L. N., 2000. Polynomial Interpolation and Clustered Grids. In: *Spectral Methods.* Philadelphia, USA: SIAM, pp. 41-50.

Verma, A., Rajput, N. & Subramaniam, L., 2003. *Using viseme based acoustic models for speech driven lip synthesis.* s.l., s.n.

Visser, M., Poel, M. & Nijholt, A., 1999. *Classifying visemes for automatic lipreading.* Plzen, s.n.

Walden, B. E. et al., 1981. Some Effects of Training on Speech Recognition by. *Speech, Language, and Hearing Research*, Volume 24, pp. 207-216.

Walden, B. E. et al., 1977. Effects of training on the visual recognition of consonants. *Speech and Hearing Research*, 20(1), pp. 130-145.

Wark, . T. & Sridharan, S., 1998. *A Synthetic Approach to Automatic Lip Feature Extraction for Speaker Identification.* Seattle, USA, International Conference on Acoustics, Speech and Signal Processing (ICASSP'98).

Waters, K., 1987. *A Muscle Model for Animating Three-Dimensional Facial Expression.* Anaheim, ACM SIGGRAPH.

Waters, K. & Frisbie, J., 1995. *A Coordinated Muscle Model for Speech Animation.* Quebec, Canada, Graphics Interface.

Wilhelms, J. & Van Gelder, A., 1997. *Anatomically Based Modelling.* Los Angeles, CA, USA, ACM SIGGRAPH.

Williams, J. J. & Katsaggelos, A. K., 2002. An HMM-Based Speech-to-Video Synthesizer. *IEEE Transactions on Neural Networks*, 13(4), pp. 900-915.

Williams, J., Rutledge, J., Garstecki, D. & Katsaggelos, A., 1997. *Frame Rate and Viseme Analysis for Multimedia Applications.* New Jersey, s.n.

Williams, L., 1990. *Performance-driven facial animation.* Dallas, USA, ACM.

Woodward, M. F. & Barber, C. G., 1960. Phoneme perception in lipreading. *Speech, Language and Hearing Research*, 3(22), p. 212.

Wu, Y., Magnenat-Thalman, N. & Thalmann, D., 1994. *A plastic-visco-elastic model for wrinkles in facial animation and skin aging.* Beijing, China, Pacific Graphics.

Xie, L. & Liu, Z. Q., 2006. *Speech Animation Using Coupled Hidden Markov Models.* Hong Kong, IEEE, pp. 600-607.

Xie, L. & Liu, Z. Q., 2007. Realistic Mouth-Synching For Speech-Driven Talking Face Using Articulatory Modeling. *IEEE Transaction on Multimedia*, 9(3), p. 500–510.

Yuille, A., Hallinan, P. & Cohen, D., 1992. Features extraction from faces using deformable templates. *International Journal of Computer Vision*, 8(2), pp. 99-111.

Zhanyu, M. & Leijon, A., 2008. *A probabilistic principal component analysis based hidden markov model for audio-visual speech recognition.* Pacific Grove, s.n.

Zhao, G., Barnard, M. & Pietikainen, M., 2009. Lipreading with local spatialtemporal descriptors. *IEEE TRANSACTIONS ON MULTIMEDIA*, 11(7), p. 1254–1265.