INVESTIGATING THE EFFECTS OF ACCENT ON VISUAL SPEECH

Amy Irwin

Thesis submitted to the University of Nottingham for the degree of Doctor of Philosophy

September 2008

Abstract

Speechreading is a complex skill affected by both the observer's method of extracting visual speech information and talker-specific variation in speech production. This thesis focuses upon accent, a factor that can influence both an observer's viewing strategy and talker speechreadability. Auditory research demonstrates that an unfamiliar accent reduces speech intelligibility. The primary aim here was to determine whether accent type, familiarity or variation would alter *visual* speech intelligibility with consequential effects upon speechreading performance. Experiments 1 and 2 considered visual discrimination of native and non-native accented speech and the influence of non-native accent upon speechreading performance. Results indicated that observers were able to utilise visual cues for discrimination and were significantly poorer at speechreading a non-native accent. Experiments 3, 4 and 5 examined the influence of regional accent on speechreading performance. Results indicated that visual speech performance was significantly worse for Glaswegian-accented talkers than for talkers with a Nottingham accent. However, no clear advantage for accent familiarity was found. Experiment 6 examined the influence of accent type and talker variability upon speechreading performance. Accent type was consistently the dominant influence upon speechreading performance, above familiarity and variation. Experiments 7, 8, 9 and 10 examined the influence of exposure, context and repetition upon the effects of a Glaswegian accent. Here, the effect of the Glaswegian accent on talker speechreadability was reduced by context and repetition, but not removed entirely.

In conclusion, while visual accent type mostly determines visual speech intelligibility, accent familiarity mostly determines auditory speech perception.

Although spoken accent effects can be quickly reduced through exposure, no such effect was found here in the visual modality. Both context and repetition were necessary to improve the intelligibility of accented speech. This indicates a potential difference in the processing of accented speech across the two modalities and has implications for speechreading training.

Acknowledgements

I am in the unusual position of having two first supervisors who deserve acknowledgement. First, I would like to thank Dr Sharon Thomas for all her scientific guidance and extensive help through the first 2 years of my PhD. I would also like to thank Prof Deb Hall for her willingness to help me in my final year, providing support, encouragement and a healthy dose of motivation!

Prof Peter Stockwell, my supervisor from the English department, deserves thanks for his comments and administrative help at the start of my thesis.

Mike Pilling proved an excellent office mate, answering my many questions with unfailing patience and always there to lend statistical support!

I could not have conducted my research without the help and patience of my 'talking heads', to them I extend my thanks.

My husband Lee Irwin was the rock on which I could lean, always there when I needed him and often shouldering an unequal share of the housework and upkeep of our horses, earning my grateful thanks.

This research is dedicated to Lee Irwin, Janet Nelson, Rye, Noggin and the Buns.

Declaration

This thesis is the candidates own original work and has not, whether in the same or different form, been submitted to this, or any other University for a degree. All experiments were designed and analysed by the candidate, and all testing was conducted by the candidate.

Conference Presentations

Irwin, A., Thomas, S. & Pilling, M. (2006, 02/05/2006) Identification of Language and Accent through Visual Speech, Speech Prosody 2006, Dresden.
Poster Presentation.

Irwin, A., Thomas, S. & Pilling, M. (2007, 31/08/07) Regional Accent Familiarity and Speechreading Performance, International Conference on Auditory-Visual Speech Processing 2007, Hilvarenbeek.
Poster Presentation.

Irwin, A., Thomas, S. & Pilling, M. (2008, 02/04/2008), British Regional Accent Familiarity and Speechreading Performance, British Psychological Society Annual Conference 2008, Dublin.
Poster Presentation.

CONTENTS

Chapter 1. Speechreading in the Literature

'But it must be admitted that the Good Lord has created few people with legible countenances' (Calkins, 1924).

Speechreading entails processing the visual signal provided by the talker's moving face in order to comprehend his/her speech. It is a complex and sometimes difficult process. Part of that difficulty stems from the low visibility of some articulatory speech movements, particularly those which are produced at the back of the mouth (Lesner, 1988). Another important influence is the high level of variability found across talkers in terms of visible speech production (Kricos & Lesner, 1982; 1985), which has been shown by several authors (Demorest & Bernstein, 1992; Lesner, 1988) to affect speechreading ability. This variability is thought to be caused by several factors, including speech rate (Massaro, Cohen & Gesi, 1993), lip shape (Lesner, 1988) and facial expressivity (Jacobs, 1982). There are two views on how talker variability influences speechreading ability, the first focuses on the method by which an observer extracts information from the visual signal (Conrey & Gold, 2006), the second focuses upon the influence of various talker factors on the intelligibility of the visual signal produced (Yakel, Rosenblum & Fortier, 2000).

One factor that may influence both talker and observer (as illustrated in Figure 1.1), but has yet to be researched in detail, is accent variation. This leads us to the overall aim of this thesis, which is to examine the effect of accent variation on the intelligibility of the visual signal and consequently upon speechreading performance. Figure 1.1 represents a detailed framework for speechreading performance, one that clearly highlights the focus of my research.

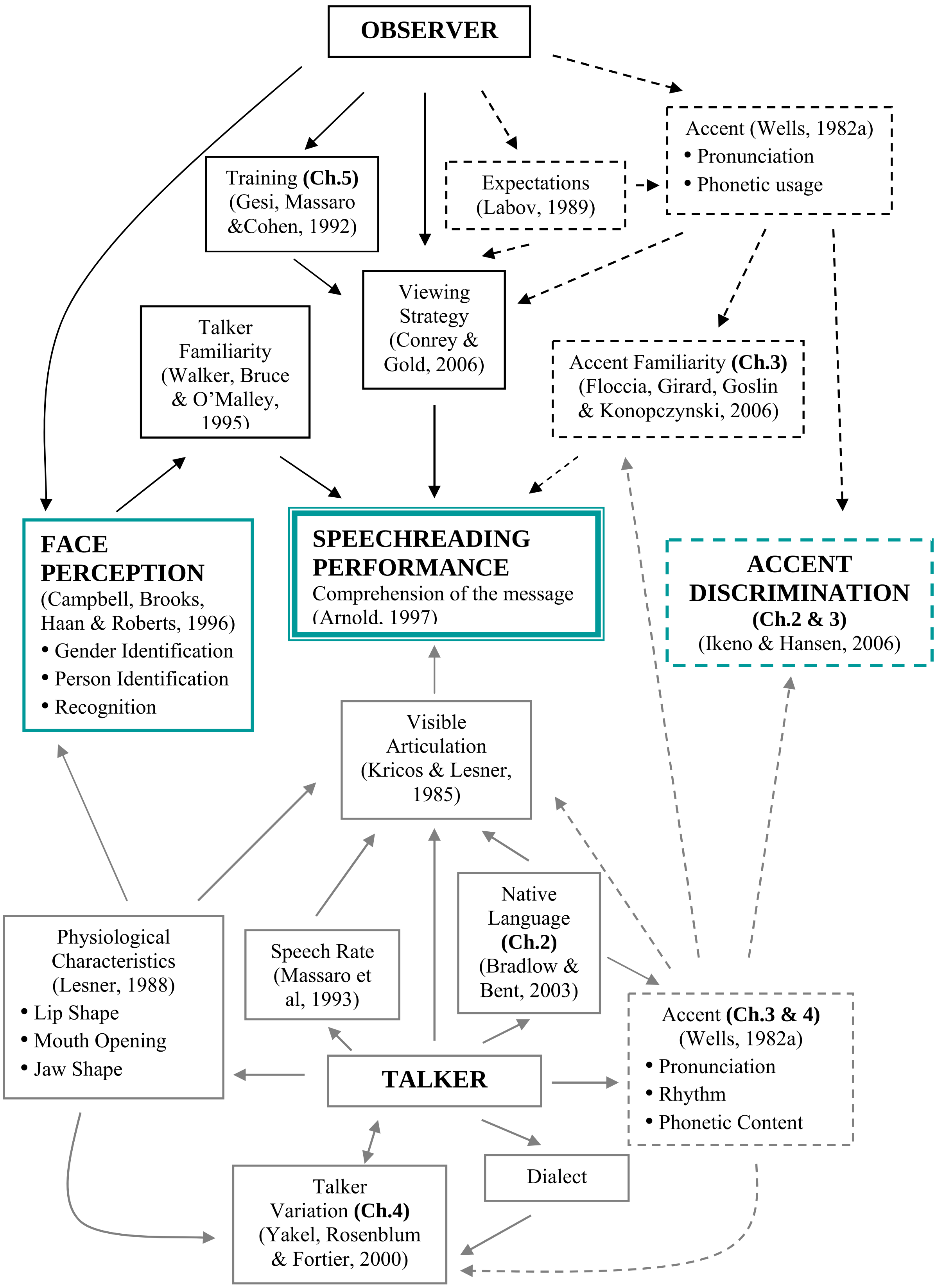


Figure 1.1: Detailed framework showing how speechreading performance (centre) is determined by observer (top, black lines) and talker (bottom, grey lines) characteristics

Those aspects of Figure 1.1 shown by a dotted line (boxes and arrows) represent factors known to influence auditory speech comprehension, but not yet investigated for visual speech perception. The grey boxes and arrows, in the lower half of Figure 1.1, relate to those aspects of speechreading performance that are determined by the characteristics of the talker. Those boxes which are outlined in black, in the upper half of Figure 1.1, are associated with those aspects of speechreading performance that are determined by the observer. Finally, those boxes outlined in green, in the centre of Figure 1.1, represent the final stages of processing; speechreading, face perception and accent discrimination. The model shown in Figure 1.1 represents the synthesis of several published papers on the subject of speech perception, bound together by a novel framework, with general inspiration from both the Bruce and Young (1986) model of face processing and Ellis's (1986) model of face recognition.

The primary aim of my thesis is to determine the degree to which the visual speech perception system operates in a similar manner to that of the auditory speech perception system. Of particular interest is whether accent variation yields a significant effect upon visual speech intelligibility, and consequently speechreading performance, in the same way that it does for auditory speech understanding. Experiments examine whether the effects of accent variation are attributable to both talker and observer; the observer in that their familiarity with an accent type should influence their expectations regarding speech production and thus their comprehension of speech, and the talker in that accent impacts on several aspects of speech production, each of which have been shown to influence auditory speech intelligibility in a variety of ways (Floccia et al, 2006).

The theoretical background to Figure 1.1 will be discussed within this chapter. The chapters that follow have each been designed to investigate hypotheses arising at different stages of the framework. Chapter 2 deals with non-native speech production, looking at both foreign accent discrimination and the effect of foreign accent upon visual speech intelligibility. Chapter 3 deals with regional accent variation; looking firstly at the opinions of deaf speechreaders regarding accent effects, then investigating regional accent discrimination and the effects of regional accent upon visual speech intelligibility. Chapter 4 deals with talker variability, comparing the effects of general talker variability with accent variation. Chapter 5 deals with the ability of an observer to adapt to an unfamiliar, or difficult, accent type, looking at the effects of context, repetition and exposure.

1.1 Visual speech perception

The production of speech involves an individual's control of their various articulators (lips, tongue, larynx etc.) to produce bi-modal (auditory and visual) speech signals (Jiang, Alwan, Keating, Auer & Bernstein, 2002). The acoustic and visual aspects of spoken language originate in the control of the passage of air through the mouth and throat, using the various articulators (Carr, 1999). Thus, as a person speaks, they pass air up through their vocal cords, the oral and nasal cavities and past their lips. The majority of voiced articulations caused by this control can be identified from the auditory signal alone. However, identifying the relevant linguistic segments using the visual signal is comparatively more difficult because the resulting articulations vary in visibility. For example, those articulations that involve the lips (bilabial e.g. /b/), teeth (labio-dental e.g. /f/) or the front of the mouth (alveolar e.g. /s/) are easier to distinguish visually than articulations which involve the back of the tongue (velar e.g. /c/) or the vocal cords (voiced versus unvoiced, e.g. /s/, unvoiced versus /z/, voiced) (Carr, 1999). The result of this is that movements that produce an intelligible auditory signal are not always sufficient to produce intelligible visual speech (Jeffers & Barley, 1971). Thus, a talker may be intelligible to a normal-hearing listener but not to a deaf observer who relies on speechreading.

Despite these asymmetries across visual and auditory modalities, speechreading constitutes a dominant method of communication for approximately 9 million deaf and hearing-impaired individuals in the UK (RNID, 2008), allowing them to communicate with the predominantly normal-hearing population. This statistic would suggest that there are aspects of visual speech production that are informative for speech intelligibility. Certainly research indicates that there are aspects of both consonant and vowel production which have visual correlates (Summerfield, 1991). For some consonants, the degree of constriction of the lips and the passage of air through them produces a distinctive visual articulation (e.g. /w/ or /b/). For other consonants, the shape of the lips whilst producing the sound, or the apparent tongue movement, enables observers to correctly discern the consonant being produced (e.g. /l/ or /r/). In terms of vowel production, the majority of vowels are highly visible when produced in ideal conditions, due to the distinctive lip motions associated with each one (Summerfield, 1991). Essentially, English vowels are distinguished by tongue position, split into two categories of height and front-back positioning, and lip shape, characterised by degree of lip-rounding (Lisker & Rossi, 1992). The degree of lip-rounding provides a distinctive visual cue as to the nature of the vowel and combined with visible tongue movements this allows the majority of observers to accurately identify vowels from the visual modality.

In summary, the visual signal supplies information about the content of a linguistic message through visible articulations. Some of the speech information is lost due to articulatory motions which do not produce associated visible facial movements (e.g. velar), but there often remains sufficient information to allow utilisation of the visual signal. Confirming this is a report from Auer and Bernstein (1997) which examined the visual distinctiveness of words based on phonemic distinctiveness, word frequency and visual similarity to other words. The researchers suggest that the visual distinctiveness of a word does not rest purely upon the visible articulatory movement. Accurate word identification also depends on the structure of the lexicon. For example, the visible articulatory movements associated with /b/, /p/ and /m/ are thought to be confusable due to the similarity of lip movements required to produce each one. However, the word 'broom' is unlikely to be incorrectly identified due to this confusability, as the alternatives of 'mroom' and 'proom' are non-words and thus likely to be discarded in the event of misperception. In comparison, the word 'bat' could easily be confused with the words 'mat' and 'pat' (Auer & Bernstein, 1997). On this basis, the study used computational modelling to estimate that 54 – 61% of words are visually unique and therefore, under optimal conditions, intelligible to the average speechreader.

Hearing-impaired individuals have been shown to utilise the information available through visual speech as a means to accentuate degraded speech signals by compensating for lost acoustic information, thus improving speech intelligibility (Fitzer, 2003). Certainly, a study designed to improve the perception and production of speech by hearing-impaired children (Massaro & Light, 2004) found that highlighting the vocal tract and visible articulators, by drawing the children's attention to them (using the computer animated head 'Baldi'), facilitated their learning. This result illustrates the use of the visual signal to augment a distorted or degraded auditory signal. In summary, it would appear that the visual signal can be used to comprehend or accentuate speech, though the success that is achieved will vary depending on factors associated with the talker, the observer and the environment in which speechreading takes place.

1.2 Audiovisual speech perception

Many studies have examined visual speech in conjunction with auditory speech in order to investigate the apparent complementarity between the two modalities in facilitating speech perception (Summerfield, 1987). Studies focus on the contribution of visual speech when the auditory signal is degraded in some way, on the contribution of visual speech to language learning and on the contribution of visual speech to everyday speech perception using illusory conjunctions. In one highly influential study, Sumby and Pollack (1954) found that presenting a talker's face improved perception of words in noise significantly, as shown in Figure 1.2.

Figure 1.2 illustrates that audiovisual presentation of words (right panel) increases speech intelligibility under degraded conditions compared with auditory presentation alone (left panel). The clear difference in performance levels between the auditory and audiovisual conditions is indicative of the relative contribution of the visual signal to speech perception (Sumby & Pollack, 1954).

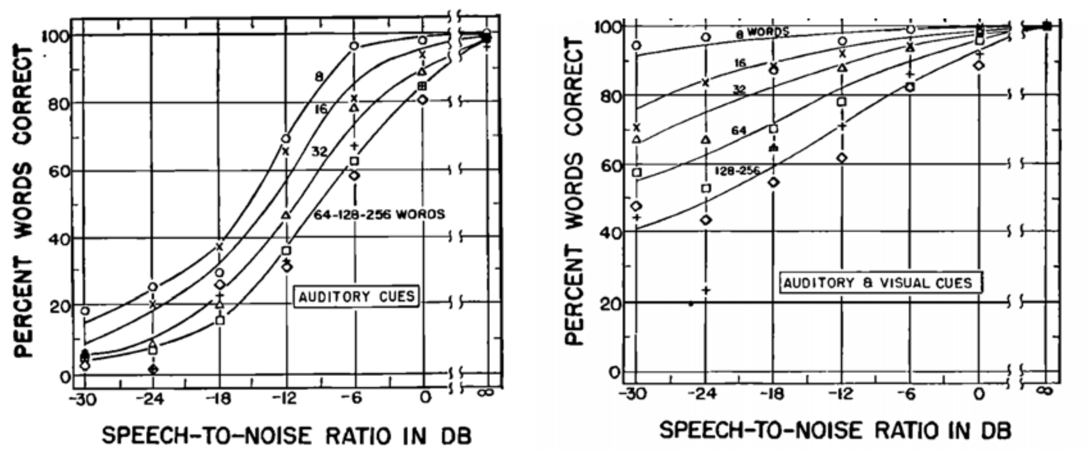


FIG. 1. Speech intelligibility under conditions of auditory presentation alone as a function of the speech-to-noise ratio under test. The parameter on the curves is the size of the vocabulary (spondee words) under examination. Each point in Figs. 1 and 2 represents the average results for 450 determinations pooled over subjects.

FIG. 2. Speech intelligibility under conditions of simultaneous auditory presentation and visual observation of a speaker's facial movements as a function of the speech-to-noise ratio under test. The parameter on the curves is the size of the vocabulary (spondee words) under examination.

Figure 1.2: Comparison of speech intelligibility (words correct) across auditory (left panel) and audiovisual (right panel) presentations (Sumby & Pollack, 1954)

Figure 1.2 also indicates that the relative improvement in speech perception offered by visual speech increases as the signal-to-noise ratio decreases. For example, at an SNR ratio of -30 dB, the improvement offered by visual speech is between 30 and 75%, at a SNR ratio of -6 dB, the improvement produced by visual speech is approximately 20%. Thus the importance of the visual signal increases as the acoustic signal becomes more degraded. The authors' state that this result highlights the importance of the visual modality in speech perception and suggest that speech intelligibility could be improved in many practical situations (such as in factories or in military situations) by ensuring that the face of the talker can be viewed (Sumby & Pollack, 1954).

Further research in this area indicates that the visual signal provides clear information relating to place of articulation; the distinction between the bi-labial /p/ and labio-dental /f/ for example, whereas the auditory signal provides less distinct information which could easily be masked by noise (Summerfield, 1987). Those aspects of the auditory signal which are less easily masked, such as the distinction between /d/ (voiced) and /n/ (nasal) are often associated with confusable visual signals. Thus, the visual signal appears to improve the perception of those aspects of auditory speech which are most susceptible to masking and vice versa (Summerfield, 1987).

Research has also investigated the effect of visual speech information on informational masking; the masking of a speech signal using one or more competing speech signals (Helfer & Freyman, 2005). The authors found that the visual signal reduced the detrimental effect of competing voices upon speech perception. The authors suggested that visual cues reduce uncertainty about when a talker begins to speak and aid in the identification of the target talker, thus increasing the focus of attention on the relevant speech information.

The studies described above concentrate on the effect of visual speech when the auditory signal is degraded. An alternative approach is to examine the interaction of visual and auditory speech when the auditory signal is intact, i.e. presented in quiet listening conditions (Arnold & Hill, 2001; Reisberg, McLean & Goldfield, 1987). Such research argues that the perception of speech is amodal in nature, with visual cues always being utilised not just when the acoustic signal is distorted or degraded (Reisberg et al, 1987). In order to investigate the potential advantages of audiovisual speech perception, the studies reduced auditory speech intelligibility through the use of semantically complex sentence materials (e.g.

sentences which reference a subject topic which a listener has no experience with, generating many unfamiliar words and phrases) or by using a talker with a strong unfamiliar accent (Glaswegian) (Arnold & Hill, 2001). Speech comprehension was then compared across auditory and audiovisual presentation of the same stimuli, with the finding that audiovisual presentation significantly improved comprehension levels. The authors suggest that speech processing integrates visual and auditory information at an early level, utilising the visual signal even when the auditory signal is clearly audible.

This suggestion is further supported by developmental studies with congenitally blind children which indicate that acquisition of speech is affected by a lack of visual information (Mills, 1987). Within the study, sighted children were shown to be faster to learn those sounds with a visible articulation, such as /b/ and /f/, than those without e.g. /d/ and /n/. The blind children were shown to be slower in the acquisition of those sounds and thus produced a different configuration of speech errors (common confusions among labial, velar and alveolar stops) (Mills, 1987). The authors suggest that the information supplied by the visual signal regarding place of articulation is a necessary part of speech production and one which influences the development of speech skills. Essentially, the lack of visual information increased the time needed to learn certain speech sounds, though it should be noted that those sounds were all eventually acquired. This indicates that while visual speech is important in the early stages of phonetic development, it is possible, in the longer term, to learn speech sounds without it.

Further evidence that visual information plays an important part in everyday speech perception is illustrated by the "McGurk" effect (McGurk & MacDonald, 1976). This effect is due to the integration of visual speech into speech perception and is automatic for most individuals. The study presented participants with incongruent auditory and visual speech tokens. For example, a combination of the auditory token /ba/ was presented in conjunction with the visual token /ga/, or the visual token /bi/ was produced with the auditory token /gi/. The results indicated that the incongruent signals could be perceived as a combination of the two signals with aspects of both the modalities included. For example, in the second case listed above the reported percept was typically /bgi/. Alternatively, a fused percept could be reported. Thus in the first case of /ba/ and /ga/, participants typically reported /da/, a fusion of both the visible place of articulation and the acoustic voicing cue (Colin, Radeau, Soquet, Demolin, Colin & Deltenre, 2002). Interestingly, this effect remained even when participants had been informed of the potential incongruence between the two sets of information (Summerfield & McGrath, 1984).

The McGurk effect has also been shown to be highly resistant to experimental manipulation, be that a reduction of the human face to a dynamic point-light display (Rosenblum, Johnson & Saldana, 1996) or the removal of colour from a face (Jordan, McCotter & Thomas, 2000). Similarly, the advantage of audiovisual speech perception remains unchanged despite a reduction in the size of a talker's face (Jordan & Sergeant, 2000) or a change in viewing angle (full face, three quarters and profile; Jordan & Thomas, 2001). This indicates that visual information reliably benefits speech perception in many situations. Such a robust effect argues for the automatic integration of the two modalities in speech perception.

In support of the earlier proposal that visual information plays a key role in the development of speech perception, the McGurk effect has also been shown to be present in infants (Rosenblum, Schmuckler & Johnson, 1997). Using the habituation method, infants were first shown the auditory token /va/ with the visual token /va/. Later the auditory token was changed to /da/ or /ba/ with the visual token remaining the same. Infants generalised across /va/ and /ba/ (no increase in interest in response to /ba/) but did perceive /da/ as different. Thus, the children were shown to generalise across different audiovisual stimuli based on visual information, an example of the McGurk effect (Rosenblum et al, 1997).

To summarise, visual information relating to the articulatory movements associated with speech production is automatically integrated with the corresponding auditory information. The visual signal serves to disambiguate elements of auditory speech which are acoustically confusable, thus improving speech intelligibility in general, but particularly when the auditory signal is degraded through masking by noise, other talkers or by a talker with a strong accent. The next point of interest within this thesis is the effect of natural manipulations of the speech signal, such as talker variability, on the observer's ability to extract meaningful information from visual speech.

1.3 Talker variation and theories of auditory and visual speech perception (observer viewing strategy)

The characteristics that produce talker variation encompass three types of information; i) information that signifies group membership, such as regional accent and dialect, ii) talker-specific information, such as properties of speech that relate to age, gender and vocal tract shape, iii) affective properties, including emotional state and health. These three types of information are collectively termed the 'indexical' properties of speech (Abercrombie, 1967). Early research on the perception of auditory speech proposed that the indexical properties of a talker and the phonetic content of the message were processed separately, with talker variability simply forming background 'noise' (Pisoni, 1997). This suggestion is strengthened by neurophysiological evidence which indicates that the two types of information (linguistic and indexical) are processed by distinct, and separate, areas of the brain (Levi & Pisoni, 2007). For example, research conducted using fMRI has reported distinct areas of activation for voice (indexical) and word (linguistic) processing (Stevens, 2004). Essentially, the results suggested that distinguishing between two voices utilised the right frontal-parietal area, whereas distinguishing between two words (linguistic comparison) utilised the left frontal and bilateral parietal areas (Stevens, 2004). Several similar studies indicate further physiological separation between speech perception, speaker recognition and emotional state (see Belin, Fecteau & Bedard, 2004 for a review).

The 'abstractionist' approach (Nygaard & Pisoni, 1998) to speech perception therefore considers variability in the signal caused by indexical information to be 'noise', which the speech processing system must remove in order to correctly process the linguistic content of the message (Nygaard & Pisoni, 1998). Within this approach, perceptual adaptation, or normalisation, has generally been accepted as being the process by which normal-hearing individuals compensate for the different auditory accents, dialects and speaking styles of talkers (Nygaard & Pisoni, 1998).

Normalisation in auditory speech perception is thus defined as the evaluation of talker-specific auditory properties, followed by the comparison of those properties to the observer's own speech prototypes. In order to understand different speakers, the observer must remove the variation between their own speech and that of the talker to allow comprehension of the message (Nygaard & Pisoni, 1998). Essentially, the speech perception system filters out any variability in the speech signal and then processes it in an abstract form, allowing the observer to comprehend speech produced in a myriad of different ways (Kraljic & Samuel, 2007). It has been suggested that this process of normalisation could result in an initial period of poor comprehension when an unfamiliar talker or speaking style is first encountered. This is a result of the system utilising cognitive resources in order to compare the incoming speech signal to the system's own set of prototypical speech sounds, or exemplars, to facilitate comprehension (Pisoni, 1997).

More recent research suggests an alternative to this view in which indexical information is processed in parallel with the linguistic content of speech. When a new talker is encountered, both the linguistic and indexical properties of speech are encoded by the perceptual system. Thus, indexical information associated with that talker becomes part of the internal mental representation of speech content. Since variation in indexical information is high, initial contact with a new speaker will cause a period of poor speech comprehension, whilst cognitive resources are devoted to encoding the information into memory (Nygaard & Pisoni, 1998). However, once an observer is familiar with a talker's articulatory habits, perception of that individual's speech steadily improves (Pisoni, 1997). Essentially, the 'abstractionist' theory of normalisation views talker variability as a problem that must be countered by removing variation from the speech signal. In contrast the 'encoding' theory from here onwards, suggests that the speech processing system 'learns' as a result of exposure to different talkers by encoding specific speech information and adjusting accordingly. Both the abstractionist and encoding theories assume that speech comprehension accuracy will, at least initially, be reduced when the perceptual system is faced with talker variability. However, the encoding theory has accrued the greater empirical support over recent years and so it is this theory upon which I shall now concentrate.

One minor point to note before moving on is that within the confines of everyday conversation, at least some of the adaptation to another speaker's voice arises from convergence. That is, when engaged in dynamic conversation with a talker whose mode of speech production differs significantly from our own, a natural response is to mediate our own speech production in an attempt to alleviate some of the variability and improve speech intelligibility. In other words, both partners in a conversation will alter their speech to make it more similar to the opposing speaker and make it easier for them to comprehend the message (Kraljic & Samuel, 2007).

Most of the research conducted on talker variability has focused on the perception of auditory speech, and has examined variation in the speech signal such as that caused by differences in the shape and length of the oral and nasal cavities, the acoustic properties of vowels (formant frequencies), voice qualities (pitch and timbre) and speech rate (Mullenix, Pisoni & Martin, 1989). Such variation in speech production has been shown to have a negative impact on the perception of vowels and consonants (Verbrugge, Strange, Shankweiler & Edman, 1976) and on the intelligibility of words presented in noise (Mullenix et al, 1989). Essentially, in both reports, it was found that varying the talker on a trial-by-trial basis impaired performance relative to a single talker. In the first study, the error rate for the recognition of vowels embedded between two consonants i.e. /p -

vowel- p/ dropped from 17% when the stimuli were produced by 15 talkers to 9% when the list was produced by a single talker (Verbrugge et al, 1976). Multiple talkers resulted in both slower and less accurate word recognition than a single talker (Mullenix et al, 1989). Together these studies emphasise the detrimental effect of multiple sets of talker-specific information upon auditory speech processing.

More recent research on the effects of multiple talkers has extended into the domain of visual speech. One such experiment evaluated the ability of normal-hearing and cochlear-implant users to recognise words presented in auditory, visual and audiovisual modalities (Kaiser, Kirk, Lachs & Pisoni, 2003). A variety of words with differing levels of lexical difficulty were included. Difficulty was based upon word frequency within the English language i.e. words found more frequently were judged to be easier than words which were infrequent. Performance was also compared across single- and multiple-talker lists in order to evaluate the effect of talker variation upon performance.

As Figure 1.3 illustrates, performance varied across the three modalities of presentation within the study, with audiovisually presented words being associated with the most accurate recognition rate. The results showed that a single-talker and lexically easy words were associated with more accurate word recognition scores irrespective of presentation modalities. Interestingly, the single-talker list produced the greatest advantage for lexically difficult words within the audiovisual condition. The authors suggest that the encoding of talker-specific information, as in the single-talker lists, should aid in the disambiguation of multiple word choices within the lexicon, resulting in the improved performance seen here (Kaiser et al, 2003). However, the information provided by the auditory or visual modality alone may be insufficient to allow rapid encoding of the necessary information thus reducing the effect.

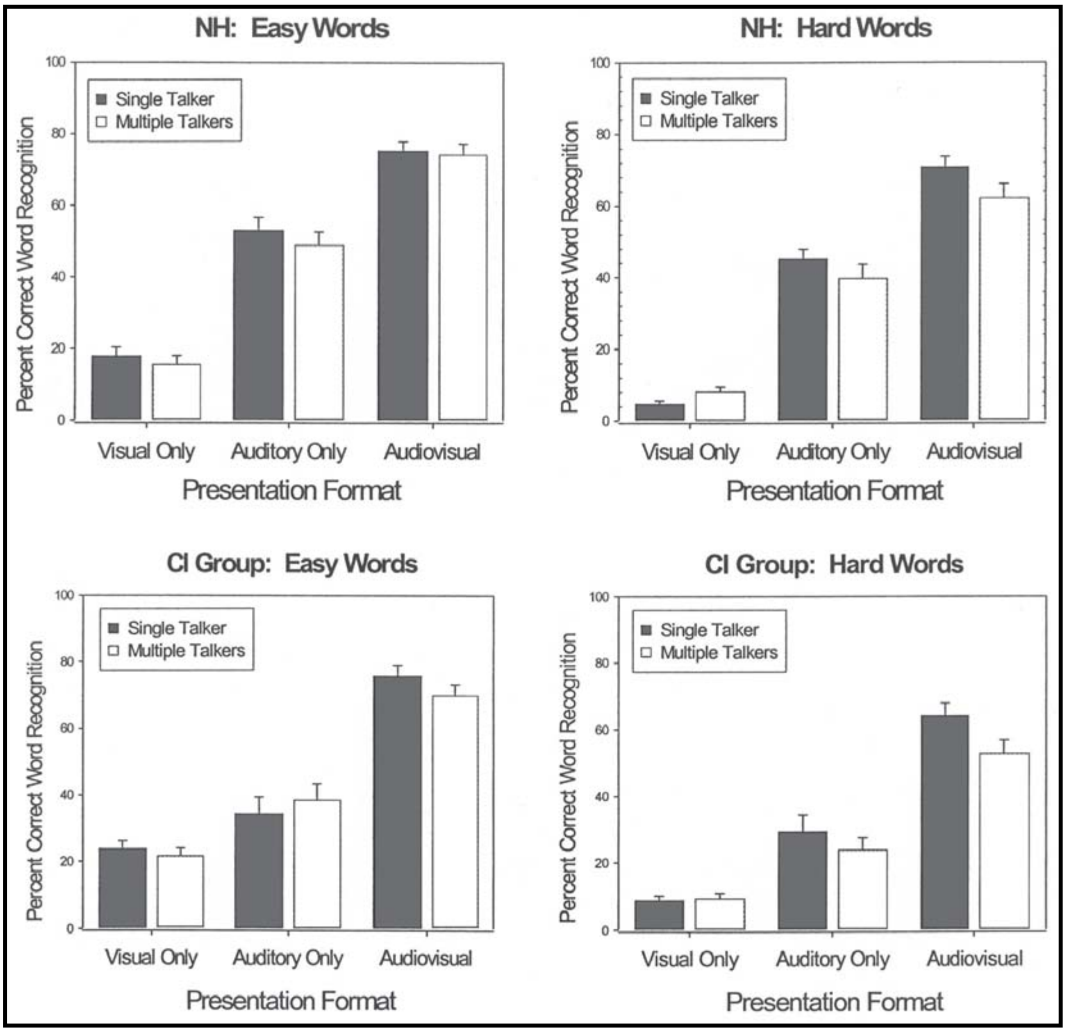


Figure 1.3: The percentage of words correctly identified in each presentation format by the normal-hearing (NH) listeners (top panel) and cochlear-implant (CI) users (bottom panel) for lexically easy (left panels) and hard (right panels) words. The parameter in each panel is the condition of talker variability (Kaiser, et al, 2003).

Application of the ‘abstractionist’ and ‘encoding’ theories to visual speech perception could involve the observer extracting all the talker-specific characteristics (i.e. skin tone, eye colour) and retaining only those relevant to linguistic content (i.e. lip movements, position of jaw) in order to normalise the incoming dynamic facial movements (Yakel, et al, 2000). Alternatively the observer could encode the talker's facial articulations and characteristics alongside the linguistic content of the message, inflicting processing costs for multiple talkers but improving the perception of a single talker. The effect of talker variability on visual speech perception has been examined by Yakel and colleagues (2000). They found that viewing multiple talkers, as opposed to a single talker, had a detrimental effect on speechreading performance. The authors suggest that this indicates a similar process of normalisation or encoding of talker information for both auditory and visual speech, although the precise mechanisms are unclear. Furthermore, the results also indicate that familiarisation, through extended viewing of one talker within their single-talker condition, may be an alternative explanation for the reported effects. Essentially, rather than the multiple-talker lists inhibiting performance, it may be that familiarity with the talker facilitated performance. However, without an appropriate baseline condition it is not possible to distinguish inhibitory from facilitatory effects.

In a later study, Rosenblum and Yakel (2001) compared the strength of the McGurk effect when participants were presented with either a single, or multiple, talkers. No significant difference was found between the two conditions, indicating that the effect of the visual signal upon speech perception was not reduced by the presentation of multiple faces and speech patterns. The authors also examined speechreading performance using simple syllables, again comparing performance between single- and multiple-talker lists. No significant difference was found. This result indicates that the processing cost inflicted when observing multiple talkers (or the familiarity advantage for a single talker) is reduced when the speech stimuli are simple and the task is relatively easy. It is interesting to note that this mirrors the result discussed previously within auditory research, where lexically easy words showed a smaller advantage than hard words for a single talker list (Kaiser et al, 2003). Thus, processing costs inflicted by using multiple talkers appear to create the greatest disadvantage when a speech perception task is relatively demanding.

The studies detailed above represent possible processing strategies for auditory and visual speech perception. Of further interest is the manner by which an observer deciphers visual speech and compensates for variation in the visual signal (Conrey & Gold, 2006). Essentially, talker characteristics can influence the intelligibility of visual speech produced and thus influence an observer's ability to speechread. However, an important question is whether this effect is consistent across all observers, that is, will all observers find a particular talker difficult to speechread or will the intelligibility of the talker also depend upon factors associated with the observer? One study that examined this question was an ideal-observer analysis of visual speech variation conducted by Conrey and Gold (2006). This analysis entailed quantifying the amount of information present within a stimulus – in this case the visual intelligibility of a talker - and defining the strategy which would result in the best performance using that information. The study then compared this 'ideal observer' with the performance of human participants in order to determine if performance was directly related to talker intelligibility or if it also varied on an observer-by-observer basis. Performance was found to vary between the human participants and the 'ideal observer', indicating that not all of the variation in performance across talkers could be explained through talker variability. The authors related these results to eye-movement data, suggesting that some observers have a tendency to focus on a talker's mouth rather than viewing the entire face and this reduced their performance. In terms of the present thesis, one hypothesis that will be examined is that an observer's knowledge of, and familiarity with, certain accent types may also influence their performance. Essentially, it is suggested that an observer's own accent (home accent) determines their prototypical expectations of visual speech, introducing the risk of misperception when a talker's visible articulation does not match those expectations.

1.4 Talker familiarity and face perception

Auditory research suggests that the specific characteristics of a talker's speech are encoded within the memory of a listener, enabling them to improve comprehension through a process of familiarisation. Evidence supporting this view includes improved word recall (Goldinger, Pisoni & Logan, 1991) and word identification (Nygaard, Sommers & Pisoni, 1994; Yonan & Sommers, 2000) when words are spoken by a familiar talker. Thus, increased exposure and hence familiarity with the indexical properties of a talker's speech, can improve subsequent perception of speech generated by that talker (Pisoni, 1997), suggesting a certain level of dependence between the processing of linguistic and indexical information.

Early theories of face perception indicated that, similar to auditory research, the three main components of face perception: face recognition, expression analysis and facial speech analysis, are all independent (Bruce & Young, 1986). Thus, there should be a functional dissociation between the processes required for recognising a face, judging a person's emotional state and understanding his/her speech (Belin et al, 2004). This is illustrated by Bruce and Young's (1986) model of face processing in Figure 1.4.

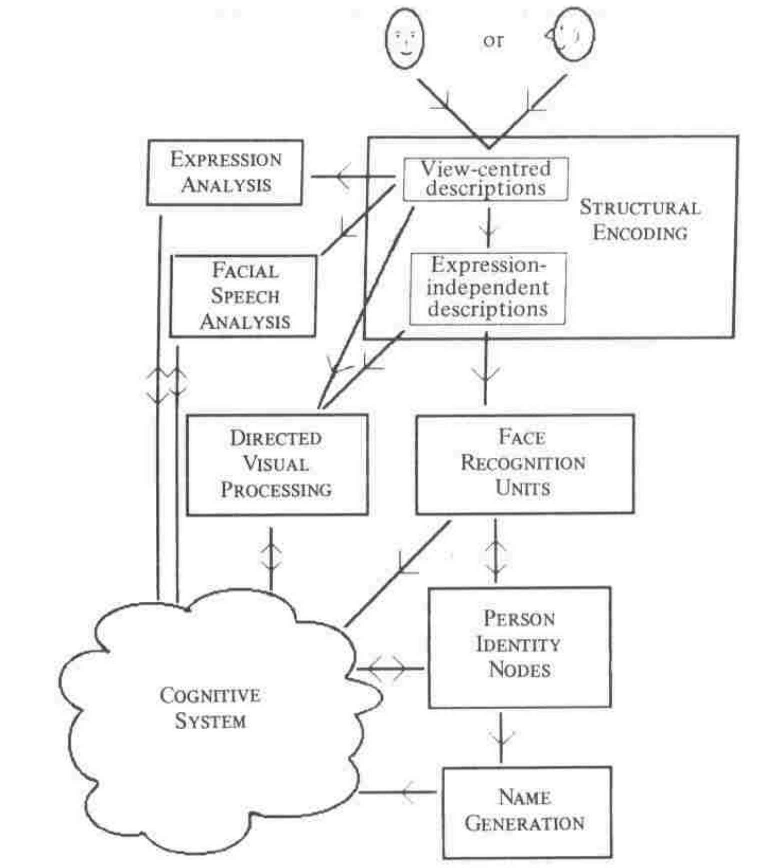


Figure 1.4: Bruce and Young's functional model for face processing (Bruce & Young, 1986)

The model devised by Bruce and Young (1986) shows a clear separation between the processes required for face perception and visual speech processing. The authors theorise that although coexisting modules perform visual speech and expression analysis, these are essentially separate from the components involved in facial recognition and consequently, facial familiarity. This implies that familiarity with a talker's facial configuration should have no direct influence upon the ability of an observer to speechread that talker. Neuropsychological evidence to support this claim indicates a double dissociation between the processes required for successful speechreading and those required for face recognition (Campbell, Landis & Regard, 1986). One patient showed impairment on facial recognition tasks but was able to speechread, while the second patient showed impairment on speechreading tasks but was able to recognise faces. However, despite such a clear indication of dissociation between the processes required for facial recognition and visual speech comprehension, more recent behavioural research in this area indicates that the dynamic aspects of a talker's face may be encoded alongside the linguistic content of the message, inflicting processing costs in multiple-talker situations (Yakel et al, 2000) and facilitating intelligibility through familiarisation. This is potentially indicative of partially, rather than fully, segregated processes for the various aspects of the visual signal.

Walker and colleagues (1995) investigated the effect of talker familiarity and gender on susceptibility to the McGurk effect. The study examined the McGurk effect using congruent and incongruent gender stimuli. There were three conditions, i) congruent stimuli (voice and face match), ii) incongruent same gender stimuli (e.g. voice and face both female but taken from different talkers) and iii) incongruent different gender stimuli (e.g. female face and male voice). Two groups of participants were recruited, one group familiar with the talkers and the other not. Those participants who were familiar with the talkers were less likely to exhibit the McGurk effect when the stimuli were incongruent than those participants unfamiliar with the talkers (Walker et al, 1995). Thus talker familiarity may produce expectations about speech production and the associated facial cues, leading to the heightened perception of incongruity in voice and face cues and hence fewer combination or fusion errors. This indicates that aspects of facial identity and speechreading may not be entirely independent.

This result can be contrasted with that of an earlier experiment (Green, Kuhl, Meltzoff & Stevens, 1991) where incongruent gender stimuli (male voice with a female face and vice versa) exerted no effect on the strength of the observed McGurk effect. The researchers concluded that the mechanism for integrating linguistic information across the modalities was not disrupted by the information being incongruous. Hence, it may be the case that talker familiarity is necessary for aspects of face processing to modulate speechreading performance. Without familiarity to modify an observer's expectations, the processing of visual speech should initially rely upon the talker's actual visible articulation alone. It might be that continued exposure to a talker is required before familiarity, and thus facial identity, may influence performance.

Further evidence for the interaction between face processing and visual speech comes from research on the impact of audiovisual speech on learning (Sheffert & Olsen, 2004). In this study, the researchers examined how visual and auditory information interact when learning to identify different talker's voices. Participants who had been trained to recognise voices using audiovisual stimuli learned at a faster rate and were more accurate at identifying voices from a new set of auditory words than participants who had been trained using only auditory information. Both groups were also tested on long-term memory for auditory words, with the words produced by both familiar (from their training) and unfamiliar talkers. The results indicated that talker familiarity was associated with enhanced word recollection, as illustrated by Figure 1.5. Performance, measured by recording the proportion of words correctly recognised as 'old' (previously presented), was significantly higher when the word was produced in a familiar voice as opposed to an unfamiliar one (Sheffert & Olsen, 2004).

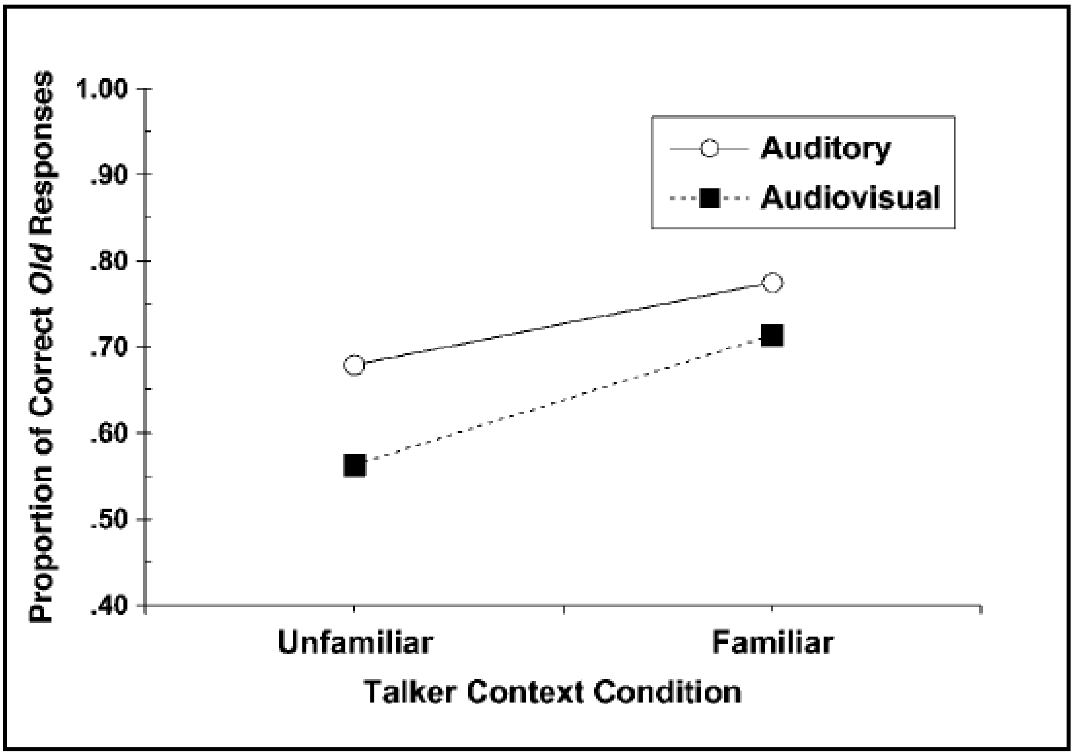


Figure 1.5: Mean proportions of correct 'old' (previously reported) responses on a word recognition task, spoken by familiar and unfamiliar talkers (Sheffert & Olsen, 2004)

The authors suggest that visual speech can provide information about the talker that is compatible with auditory talker-specific features, speeding the learning process. Furthermore, familiarity with a talker improves both the encoding of speech information and its recollection as exhibited by the improved word recollection scores.

Finally, anecdotal evidence provided by deaf speechreaders has previously indicated that familiarity with a talker leads to a greater ability to decipher that talker's speech, even if their face is slightly obscured (Cohen, 1995). The research described here further illustrates the importance of familiarity with a talker, with auditory and visual research indicating an advantage for both speech intelligibility and word recall.

1.5 Talker variation: Physiological factors, speech rate and visible articulation

A speechreader is likely to encounter thousands of different talkers throughout their lifetime, each with their own distinct speech patterns. Among these individuals will be those who are visually articulate and those who are visually incomprehensible (Lesner, 1988). Several factors could have an impact on the clarity of the visual signal produced by a talker. These include, lip shape, speed and rhythm of speech production, accent and so on. Each of these factors influences the dynamic movement of the face, altering the talker's visible articulation. Rosenblum, Smith, Nichols, Hale and Lee (2006) used point-light displays of facial movement to examine cross-modal talker matching (i.e. identifying and then matching a talker's voice and face). Participants were able to perform the task at a level significantly above chance. Thus, every talker has an identifiable speaking style that markedly affects both auditory and visual aspects of their speech production. If every talker's mode of speech production is distinctive, this implies that the production of speech varies across talkers. An observer must, therefore, be capable of adapting to different speaking styles in order to comprehend visual speech produced by different talkers.

Talker differences are sufficiently important that they should be carefully considered by researchers developing a test or training programme for visual speech. If certain talkers are easier to speechread than others this introduces variability into a test situation, both in terms of speech intelligibility and possibly also a subject-by-talker interaction (Demorest & Bernstein, 1992). Thus, specific talkers may be less intelligible than others to only a proportion of observers, the reminder finding the same talker highly intelligible. The consequence of this is that the scores generated by using different talkers are not necessarily equivalent, lowering the likelihood of comparative results across talkers and observers from different tests of speechreading ability (Demorest & Bernstein, 1992). Given these potential confounds it is important to investigate in more detail the sources and effects of talker factors upon the intelligibility of the visual signal.

Research on visual differences across talkers has been conducted by Kricos and Lesner (1982) through utilising comparisons of viseme production. A viseme is a basic unit of visual speech and usually encompasses several phonemes that have similar visual articulations (Kricos & Lesner, 1982). Thus a viseme may be distinguished from another viseme, but the phonemes within a viseme group are usually indistinguishable. Past research has theorised that there are twelve basic viseme groups, as illustrated by Figure 1.6.

/p, b, m/
/f, v/
/w/
/r/
/θ, ð/
/t, d, n/
/l/
/s, z/
/ʃ, ʒ, tʃ, dʒ/
/j/
/k, g, ŋ/
/h/

Figure 1.6: Viseme groupings (Kricos & Lesner, 1982)

The production of individual visemes and short sentences were compared across six talkers, all of whom were female and American (Kricos & Lesner, 1982). Cluster analysis was conducted on the pattern of errors recorded for each talker, with a resultant set of viseme categories constructed for each individual. Unlike previous research which had theorised that the 12 viseme groups should apply to the majority of talkers, the results reported by Krisos and Lesner (1982) found that viseme categories varied across the talkers. For example, one talker produced only four viseme categories, with one single category containing /t, d, s, z, n, l, j, h/. In contrast, another talker produced eight viseme categories, with one example containing only /t, d, s, z/. In addition, those talkers who were visually intelligible were also those who produced the largest number of viseme categories. The result was further supported by a second study that asked a group of hearing-impaired teenagers to speechread two of the six talkers used previously (Kricos & Lesner, 1985). One talker had previously been categorised as intelligible (large number of viseme categories) the other as difficult to speechread (low number of viseme categories) (Kricos & Lesner, 1985). As before, the talker who produced the larger number of viseme categories was more intelligible. Both studies by Kricos and Lesner (1982; 1985) suggest that visual speech intelligibility varies from talker to talker, even when, in this case, all of the talkers were of the same gender and general accent type.

Several factors could influence the clarity of the visual signal, causing the effects reported by Kricos and Lesner (1982). The first factor to be discussed here is speech rate. It has been suggested that the typical speech rate is too fast to be optimal for speechreading (Berger, 1972). Certainly, as speech rate increases it is more likely that errors in articulation will occur, causing a possible distortion in the visual signal (Lesner, 1988). An increase in speech rate is also likely to increase co-articulation (changes in articulation depending on the preceding, or upcoming, speech sounds, Cohen & Massaro, 1993), blurring the boundaries between words. Thus individuals who speak quickly are likely to be difficult to speechread. Research conducted by Massaro and colleagues (1993) found that presenting words at a rate that was three times faster than the average speech rate did hamper speechreading performance. Increasing the speed of speech production reduces the time allowed for perceptual processing, leading to increased error rates. Therefore a slow rate of speech should be expected to increase speechreading accuracy. However, results pertaining to this have been mixed, with several studies recording an advantage for slowed speech, whilst others have found no difference in performance (Berger, 1972). Thus, although a fast speaking rate would appear to make speechreading more difficult, it cannot be conclusively stated that a slower rate of speech would provide an advantage, since other factors also determine the clarity of visual speech.

For example, lip shape may be important since talkers with thicker lips are reported as having a detrimental effect on speechreading performance (Berger, 1972). Other factors include, facial configuration, teeth visibility (Summerfield, 1992) facial hair, enunciation, accent type and lip movements (Lesner, 1988). All of these factors contribute to talker variability and thus could influence speechreading performance. However, although each of these factors is generally accepted as having an influence, the relative strength of each one is unknown. As such it is not yet possible to predict, with accuracy, a talker's clarity of visual speech production. Further research is therefore required in this area.

1.6 Language discrimination and non-native accent

The research presented within this thesis focuses on accent as a factor of talker variability, encompassing both non-native and regional accent variation. Although regional accent variation produces changes in the pronunciation of speech, there remains an underlying consistency between talkers in terms of the phonology of a language. Speech production by a non-native talker can produce variation on a wider scale than that encompassed by regional variation, including the phonology of their first language (Bent & Bradlow, 2003). For example, English speakers consistently produce the French vowel [y] as [u] and Portugese speakers produce the same vowel as [i] (Rochet, 1995). For this reason, it is sometimes the case that non-native speakers of a language find that language easier to comprehend when it is produced in their own native accent (Bent & Bradlow, 2003). This benefit appears to depend upon the fluency of the listener. For example, research into auditory perception of second languages (L2) presented in noise, indicated that speech-reception-threshold (SRT) scores increased when a non-native language was used, as illustrated by Figure 1.7 (Wijingaarden, Steeneken & Hontgast, 2002). The SRT scores represent the speech to noise ratio at which 50% of the stimuli are perceived correctly. Thus, the recorded increase in SRT indicates that speech intelligibility was reduced through non-native production, consequently the decibel (dB) level of masking noise was reduced to increase the ratio of speech to noise and improve speech intelligibility.

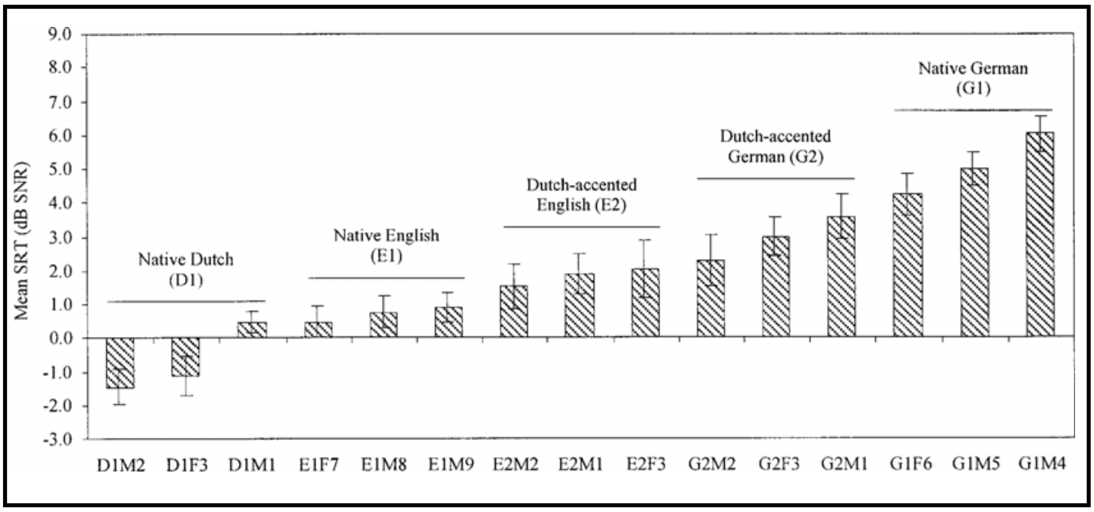


Figure 1.7: Mean SRT results of subject group I per individual talker (*N*59). All listeners were Dutch students, speaking English as a second language and German as a third language. Speech material was in Dutch (D1), English (E1 and E2), and German (G1 and G2). Non-native talkers (E2 and G2) were all Dutch (Wijingaarden et al, 2002).

Of particular interest within Figure 1.7 is that the listeners, who were native Dutch speakers with fluent English, showed a slight increase in SRT when English was spoken in a Dutch accent in comparison to English spoken using an English accent. In other words, the use of their own familiar accent actually reduced speech intelligibility (Wijingaarden et al, 2002). The opposite effect was shown for German – a language the listeners had only basic proficiency in. The researchers suggest that fluency in a language produces a set of expectations regarding the production of that language. The pronunciation of that language in the listener's native accent disrupts speech perception because although the accent type is familiar, it is not appropriate for the language and the subsequent alteration to the speech signal therefore reduces speech intelligibility. The opposite is true where fluency is not present, as with the German language, because in this case the familiar accent actually aids speech recognition by mapping the language on to a familiar set of speech expectations. It would therefore appear that production of a language in a non-native accent can have a detrimental or an advantageous effect, depending on the relative experience of the listener.

These differences relate not only to the auditory modality but to the visual modality as well. Languages differ in the weight given to visual cues, leading some observers to direct their attention primarily to the auditory signal (Hazan, Sennema, Faulkner, Ortega-Llebaria, Iba & Chung, 2006). This is particularly true for the more tonal languages such as Cantonese, where the visual signal holds less value (Hazan et al. 2006). As such, the use of the visual signal will differ between cultures. Learners of a second language must therefore attune to the appropriate visual cues in order to procure a benefit from the visual signal when they attempt to comprehend and produce speech.

Research has also examined language discrimination using visual speech. The first of these studies examined Spanish and Catalan participants' ability to discriminate between two Catalan phonemes using auditory, visual and audiovisual speech (Navarra & Soto-Faraco, 2007). The native Catalan speakers could discriminate between the phonemes using either auditory or audiovisual speech, but the native Spanish speakers were only successful in the audiovisual condition. Neither group was successful in using visual speech. Thus, although not strong enough to allow discrimination alone, visible articulatory differences between the two languages did influence perception in the audiovisual condition, improving the performance of the Spanish participants.

The same authors carried out a second study which took the discrimination of languages using visual speech one step further (Soto-Faraco, Navarra, Weikum, Vouloumanos, Sebastian-Galles & Werker, 2007). In a series of four experiments they examined participants' ability to discriminate between Catalan and Spanish sentences on the basis of visual speech alone. The results showed that bilingual speakers of both languages and mono-lingual Spanish speakers were able to discriminate between the languages. Participants from Italy and England performed at chance, these results are illustrated by Table 1.1.

Discrimination performance scores in Experiments 1-3

	Exp.1			Exp. 2		Exp. 3
	Spanish dominant	Catalan dominant	Simultaneous bilinguals	English speakers	Italian speakers	Spanish monolingual speakers
Perc. correct	57.4% * (SD=11.4)	60.9% * (SD=14.1)	59.8% * (SD=13.4)	53.1 (SD=15.8)	52.2 (SD=14.5)	55.8* (SD=11.7)
Hits / f.a.	.591 / .445	.621 / .402	.588 / .391	.528 / .466	.536 / .487	.542 / .425
d' / Crit.	.39 / .14	.63 / .28	.56 / .31	.17 / .09	.12 / .03	.32 / .20

Note. Correct discrimination percentage (Perc. correct), hits and false alarm (f.a.) rates, and d' and criterion (crit.) parameters, are presented as a function of language background in Experiments 1 through 3. The asterisk indicates significantly higher than chance scores for the discrimination percentages.

Table 1.1: Discrimination scores across three language discrimination experiments, with comparisons using the language backgrounds of participants (Soto-Faraco et al, 2007)

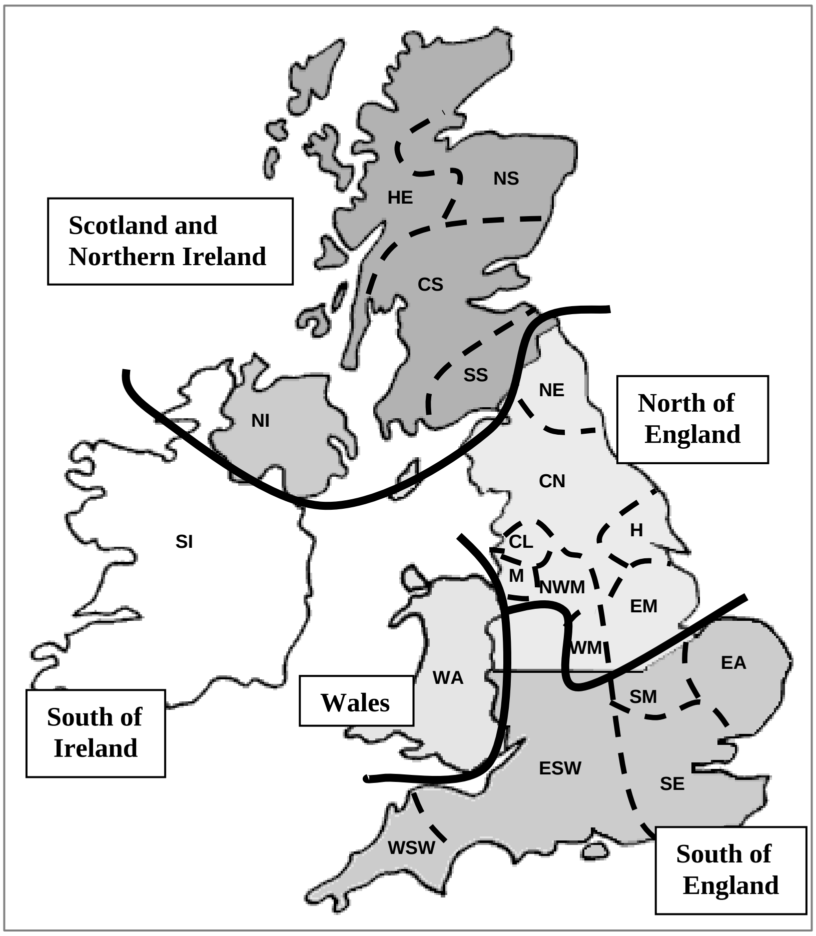
The results showed that the ability to discriminate between Spanish and Catalan using visual sentences was constrained by linguistic experience, namely that familiarity with at least one of the two languages was necessary for accurate discrimination. The authors suggest that a combination of visual cues were utilised by those observers familiar with the languages. First, observers may have been able to distinguish the subtle phonological differences between the languages using visual speech correlates. Second, the participants may have been able to determine rhythmical differences between the two languages using supra-segmental cues (word stress, vowel lengthening etc). Evidence supporting this suggestion is the finding that there are reliable visual cues that denote rhythm in speech, such as regular head nods (Munhall, Jones, Callan, Kuratate & Vatikiotis-Bateson, 2003). Finally, the participants may have used lexical cues, accessed through speechreading, to discriminate the languages. However, the ability to discriminate between the languages was not based on speechreading ability alone since the recorded speechreading performance was very low (2.5%). Thus, the visual signal appears to carry specific information denoting language type that a native speaker of that language is able to decode, without the need for a high level of speechreading ability (Soto-Faraco et al, 2007).

1.7 Regional accent

The studies described above indicate that the visual signal contains sufficient information to support language discrimination judgements, even when the differences present are subtle (Soto-Faraco et al, 2007). Although accent alters only the pronunciation of words rather than the words themselves, I predict that observers who are familiar with at least one of the accents should be able to discriminate between accents using similar visual cues. For example, like different languages, different accents could also produce changes in the segmentation of speech, rhythm, speech rate and pronunciation. However, accent variation does not change the actual content of a native language, removing some of the cues present in language discrimination and potentially making accent discrimination a more difficult task.

Regional accent is discussed here under the assumption that the language used in each case is the same it is simply the manner of pronunciation that changes (Wells, 1982a). Thus, any reference to regional variation will be referring to accent rather than dialect differences. Dialect encompasses grammar and vocabulary together with pronunciation variation (Hughes, Trudgill & Watt, 2005). For example, the sentence 'I can't go' in Standard English, becomes 'I cannae go' in the Scots dialect. Other examples include substitution of the word 'river' in the Standard English dialect with 'burn' in the Scots dialect. These differences represent only a small portion of dialect variations, hence it was decided to concentrate purely on accent variation to prevent dialect differences from becoming a confusing factor. The term 'accent' is therefore used throughout this thesis to refer solely to the pattern of pronunciation used by individual talkers, as defined by their membership of a particular regional or social group (Wells, 1982a). As such every individual produces English with an accent that is not only particular to them, but also indicative of their geographical location. Accent type is defined by the production of specific vowel and consonant sounds together with associated rhythmic, intonational and prosodic features (Wells, 1982b). In order to describe these features it will be necessary to use the phonetic alphabet together with standard lexical sets. Please refer to Appendix 1 for descriptive tables of both.

There are many differences in the production of English across Britain, particularly across regions, each of which serves to change the manner of production and sound of speech. Figure 1.8 illustrates the regional locations of some of the main accent types in the UK.



KEY

NE: North-East	WM: West Midlands	CN: Central North
SM: South Midlands	CL: Central Lancashire	ESW: Eastern South-West
M: Merseyside	WSW: Western South-West	NWM: North-West Midlands
SE: South East	EM: East Midlands	EA: East-Anglia
SI: South Ireland	NI: North Ireland	SS: Southern Scotland
CS: Central Scotland	HE: Highlands English	NS: Northern Scotland
H: Humberside		

Figure 1.8: Accent Groups in UK and Ireland (adapted from Hughes, Trudgill & Watt, 2005 & Wells, 1982). The thick black lines represent the boundaries between the five main accent regions, the dotted lines represent the boundaries between sub-regions.

It has been proposed that there are five main accent regions; the south of England, the north of England, Wales, the South of Ireland and Scotland and Northern Ireland (Hughes et al, 2005). The accent in each of these five regions has a set of general pronunciation features that are shared within that region but are distinguishable from the rest of the UK. These features allow listeners to group talkers according to their regional location. For example, in Scotland the accent is rhotic, in that /r/ is pronounced at the end of a word and after a vowel. In the South of England the accent is non-rhotic, so words such as 'bar' would be pronounced 'ba', with no emphasis on the /r/ at the end of the word (Hughes & Trudgill, 1987). The Welsh accent lacks an /h/, the North of England accent lacks the vowel /ʊ/ in putt, so the words putt and put become indistinguishable. The South of England accent lacks the distinction between /a/ and /a:/ (Hughes et al, 2005) and so on. For a full list of the main accent differences please refer to Appendix 2.

Figure 1.8 also illustrates further distinctions within each of the five main regions, these smaller sub-regions contain many of the features that define the region as a whole, but also contain differences that are sufficient for them to be segregated from one another (Hughes et al, 2005). Essentially, regional accents within the UK represent a continuum, with changes becoming more distinct the greater the geographical distance between two points. This continuum means that the ability to identify an accent type, particularly on a sub-regional level, can sometimes depend upon the level of familiarity an individual has with it. For example, a Nottingham (EM) city resident may be able to classify an Sheffield (CN) accent (located 40 miles away) as different from their own, but it is likely that an outsider would find the two accents indistinguishable and classify them both as simply ‘Midlands’, or perhaps even ‘English’ accents (Wells, 1982a). Thus, although the accent differences between the five main regions might be obvious, the distinctions within each region tend to be less distinct.

Accent types can vary from one another in four main ways: systematic differences, realisational differences, lexical-incidental differences and phonotactic differences (Wells, 1982a). Systematic differences refer to differing phonemic systems across accent types (Wells, 1982a). Essentially, accents may differ in the number of phonemes used. For example, most South of England accents include the two vowels /ʊ/ and /u/ for ‘boot’ and ‘foot’ words, whereas the Scottish accent has only a single vowel /ʏ/. Thus, at that juncture a Scottish accent has one less phoneme than an English accent. Realisational differences refer to the phonetic realisation of speech sounds. For example, two accents may differ in the relative length of a spoken vowel sound, or whether that vowel is produced as a monophthong (single pure sound, one tongue position) or a diphthong (two sounds blended together, two tongue positions). These realisational differences come in many forms and are responsible for much of the rhythmic diversity present within the accents of Britain (Wells, 1982a). Lexical-incidental differences, on the other hand, relate to the choice of phonemes within lexical items. That is, many of the pronunciation differences which allow observers to identify accent often relate to the contrasting use of phonemes between accent types. Finally, phonotactic differences relate to the distribution of phonemes within an accent, or permitted phonological structures. Essentially the phonotactic characteristics of an accent type denote the environment in which phonemes can occur (Nathan, Wells & Donlan, 1998). Thus the rhoticity of the Glaswegian accent allows /r/ to be placed in a wide variety of phonetic contexts. In comparison, the non-rhotic South of England accent restricts the phonotactic distribution of /r/ by excluding it from pre-consonantal and absolute-final (word-end) environments (Nathan et al, 1998).

The study of accent and its effect upon speech comprehension has, until now, focused on the auditory domain. These studies can shed light on potential visual correlates, with implications for the effect of accent on speechreading performance.

The auditory speech perception literature suggests that accent is a salient aspect of speech production. For example, participants are able to accurately detect pronunciation by a non-native speaker from only one example of a spoken syllable (Flege & Hillenbrand, 1984; Magen, 1998). Sensitivity to accent at a phonetic level indicates that accent is a salient speech characteristic. Furthermore, lack of familiarity with an accent has been shown to have a negative effect upon auditory speech comprehension in children (Nathan et al, 1998; Nathan & Wells, 2001) and adults (Labov, 1989; Munro & Derwing, 1995; Anderson-Hsieh & Koehler, 1988) and upon speech-in-noise intelligibility (Clopper & Bradlow, 2006).

One study compared the performance of 4 and 7 year olds on their ability to first repeat and then define words (Nathan et al, 1998). Words were pronounced using either the child's familiar 'home' accent (London – South East England) or in an unfamiliar accent (Glaswegian – Central Scotland). 'Home' accent refers to accent type used by an individual, produced by their upbringing in a particular regional location. The older children were generally better at the task, although their performance was still adversely affected by the Glaswegian accent (94% correct responses to speech in London accent, 71% correct to speech in

Glaswegian accent). The younger children found it very hard to repeat and define the words spoken with a Glaswegian accent (82% correct for speech in a London accent, 43% correct for speech in a Glaswegian accent). This suggests that the ability to process an unfamiliar accent improves with age. Interestingly, the younger children were also more likely to attempt to reproduce the Glaswegian accent when they repeated the words. The authors of the study suggest that, first, the older children were better able to comprehend accented speech due to their more developed phonological representations of speech, together with their greater exposure to accent variation. Second, the attempt at a Glaswegian accent by many of the younger children may be a sign of a greater sensitivity to phonetic detail than that possessed by the older children. This sensitivity reduces the child's ability to recognise the production of a word as an unusual rendition of a previously stored lexical item. In other words, the young children viewed each accented word as a new word, one which they were unfamiliar with. In contrast, the older children were more able to recognise the accented productions as familiar words. Finally, it is interesting to note that in a later experiment in which the task was a simplified picture–word matching task, accent effects were reduced (Nathan & Wells, 2001). Thus, when the demands of the task are low, accent may exert a weaker effect on performance

Both studies by Nathan et al (1998; 2001) suggest that as a child's vocabulary increases his/her ability to distinguish phonetic differences also increases, allowing him/her to comprehend unfamiliar accents more accurately. This effect is further described by a model of speech learning developed by Flege (1992) which states that the perception of speech consists of three stages; phonemic, phonetic and auditory processing. Flege (1992) suggests the auditory aspect of speech processing develops first, followed by categorisation of sounds at the phonetic and phonemic level. In the first instance, most perceivers will categorise sounds at the phonemic level. For example, in the English language [t^h] and [t] are phonetically different but do not indicate any differences between words, so are classified by most English speakers as the same (Flege, 1992). During language development, individuals organise speech sounds into relevant categories and form prototypes (ideal set of properties within each category). With increased experience, individuals are able to form a 'tolerance region' around prototypes that allows them to identify differences in sound production as 'foreign' and thus process those sounds in a meaningful way (Flege, 1992). Sometimes tolerance regions overlap, leading to possible misidentifications of speech sounds. Although increasing vocabulary and experience will facilitate the understanding of accented speech, errors may still occur.

Further research with adult listeners has investigated the importance of context for understanding auditory speech produced in a strong or unfamiliar accent (Labov, 1989). The study was primarily made up of a 'gating' experiment, where target words were presented initially without context and then later within a framing phrase and then finally within a sentence. Participants were therefore given more context each time they tried to identify the word. The words were produced by talkers from the Chicago area in North America, with participants recruited from three North American regions; Chicago, Birmingham and Philadelphia. Although increased context did improve accuracy, even full sentence context was not always enough to produce 100% performance (performance ranged from 50 – 90% correct for the sentence condition across all participant groups). The Chicago participants were more accurate than participants from the other two regions at all levels of the task, but the difference in performance was most notable when words were presented in a sentence context (for example, the percentage of correct responses to the word 'socks' in a sentence context was 90% correct for the Chicago participants, 65% for Birmingham participants and 60% correct for Philadelphia participants). These results were compared with participants' performance when the word was missing from a sentence, resulting in the participants using context to determine it. Participants' were actually better at the task when the word was missing than when it was spoken in an unfamiliar accent (Labov, 1989). This suggests that unfamiliar pronunciation has a greater detrimental effect on comprehension than word absence. In conclusion, unfamiliar accent is enough to impair speech intelligibility even when a word is presented in context. Thus, speech comprehension is most accurate when the talker's accent is familiar to the listener.

This claim is supported by additional research detailing the effects of French regional accent variation on the processing of (French) auditory speech (Floccia, Girard, Goslin & Konopczynski, 2006). Through a series of experiments, the effect of regional accent on speech processing was examined using a lexical decision-making task (word / non-word discrimination), with performance measured using reaction times. Each participant was presented with a carrier sentence which had a dual purpose of providing context for the word / non-word to be discriminated and exposure to the accent type of the talker. In order to generalise effects across talkers, accents and listeners, each accent was produced by a minimum of two talkers. The result of the first experiment showed that participants were slower to discriminate speech produced in an unfamiliar accent, suggesting that speech processing is made more effortful when an unfamiliar accent is encountered. This result was replicated in their second experiment which compared reaction times across words in sentences and words in isolation, as illustrated by Figure 1.9. The deficit in performance associated with an unfamiliar accent was significant in both cases.

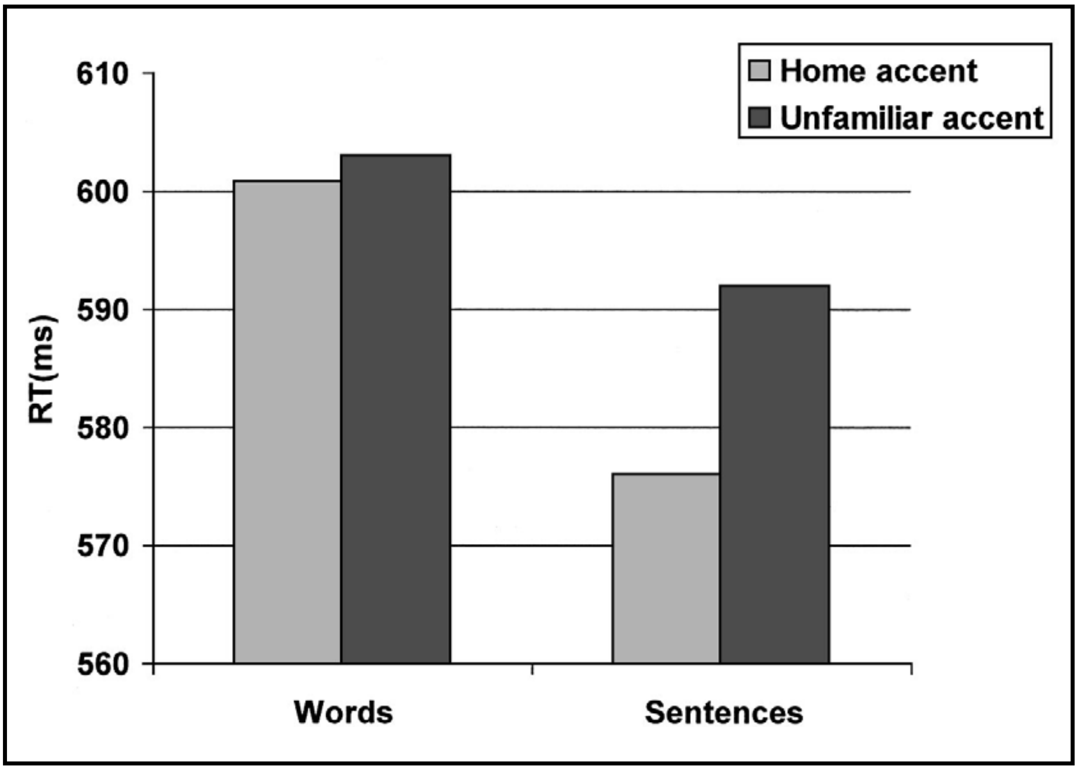


Figure 1.9: Mean reaction times as a function of target stimulus context and accent familiarity (Floccia et al, 2006)

The preliminary examination of accent effects recorded the reaction times of participants from the Franche-Comte region of France when speech was produced by talkers with a familiar Franche-Comte or Parisian accent, or the unfamiliar Toulouse or Swiss French accent. In order to determine whether the results shown above resulted from accent familiarity or a characteristic of the Toulouse accent, the authors recruited participants from Toulouse and repeated the experiment described above. The Toulouse participants were significantly slower when discriminating words / non-words when speech was produced in the Franche-Comte accent, which they were unfamiliar with, than their own familiar Toulouse accent. Thus the reaction times of the participant groups differed according to the region from which the participant came – directly linked to a participant's geographical location and their associated exposure to accents. It therefore appears that the level of exposure to an accent is important for speech processing, rather than any specific articulatory characteristics of the accent itself. If certain accents were simply more difficult to understand than others, then the pattern of results would be the same for all participants regardless of region or accent exposure (Floccia et al 2006). Thus, for every listener, there will be a distinct pattern of familiar and unfamiliar accents specific to them depending on their region of birth and their exposure to different accents.

When the same experiment was repeated with the accent variation occurring on a block-by-block basis rather than trial-by-trial basis, the effect of unfamiliar accent was significantly reduced. This suggests that repeated exposure to an unfamiliar accent improves speech intelligibility for that accent type. This appears to be an extension of the talker familiarity effects discussed previously in section 1.4, where speech produced by a single familiar talker has been shown to improve word identification (Nygaard et al, 1994). A similar experiment examined the effect of foreign accent on the speed of speech processing (Clarke & Garrett, 2004). The task was to identify whether a probe word matched the final word of an auditory sentence. One minute of exposure to a non-native accent was sufficient to improve performance to a level equivalent for speech produced in their own native English accent. These results indicate that the system used to process speech is flexible and can adapt to different accents after only short periods of exposure (Clarke & Garrett, 2004).

Thus, both talker and accent familiarity can confer an advantage for auditory speech intelligibility. In terms of visual accent variation, it could be argued that different accents produce specific differences in the movements of articulators, relating to differing stresses and pronunciations of words. Thus, each time an individual encounters a different accent, it could take them some time to become accustomed to the different facial movements and encode them into their long-term memory. This period of perceptual adjustment should be similar to that found in auditory research (i.e. Floccia et al, 2006), since, given the degree of similarity between the effect of talker variability on auditory and visual speech perception, it is possible that similarities also exist for accent variability. Thus, previous auditory research on accent can inform the visual research presented in this thesis.

Such guidance is important as unlike auditory speech perception there has been no comprehensive research to date on the effects of accent on visual speech perception. However there are indications as to the potential effect of accent on speechreading performance. First, the visual signal facilitates the comprehension of foreign speech, and, in turn, aids learners in their production and recall of a second language (Davis & Kim, 1998, 2001). This suggests that the visual signal carries meaningful information about the articulation of foreign speech that a learner is able to use, and is similar to the effect of visual speech upon talker voice identification (Sheffert & Olsen, 2004). In both cases it would appear that the visual signal enhances the auditory component of speech, improving comprehension and retention of speech information.

The second potential effect of accent on speechreading relates to the different viseme groupings shown by Kricos and Lesner (1982). Regional accent differences alter the phonetic realisation of speech sounds, and as such are likely to also affect viseme distinctiveness. Research has already shown that viseme groupings can affect the intelligibility of the visual signal (Kricos & Lesner, 1982), alterations produced by accent type are also likely to have such an effect upon intelligibility. The resultant effect of this upon speechreading performance is currently unknown.

The third potential effect comes from a study by Ellis, MacSweeney, Dodd and Campbell, (2001) that developed a new measure of speechreading. The authors noted that deaf participants from the North of England were significantly poorer than those from the South of England at matching pictures to visual speech. Poor performance by the North of English participants may have resulted from the group's lack of familiarity with the southern accent of the two talkers used in the study. This result was secondary to the study's main aim, but it does indicate that accent differences can alter the intelligibility of visual speech. At this point it is unclear whether this result is simply due to lack of familiarity with the accent type or characteristics associated with the accent type itself. This relates back to the effect of observer accent upon viewing strategy discussed in section 1.4. Essentially, it is possible that an observer's knowledge of their 'home' (familiar) accent may influence their perception of visual speech through the generation of expectations relating to visible articulation. These expectations may lead to confusion when the talker fails to produce the expected visual facial movements. Consequently, the speechreading performance of an observer will decrease, potentially leading to the results reported by Ellis and colleagues (2001). This hypothesis will be examined in detail within this thesis in Chapters 2 and 3.

1.8 Speechreading training

Various factors influence speechreading performance, including talker intelligibility and observer viewing strategy (Conrey & Gold, 2006). Individual differences in speechreading ability have been examined by several studies, with performance found to be extremely variable across participants (Jeffers & Barley, 1971). A more recent study estimated that factors associated with the observer account for 10.5% of the variability in performance (The remaining sources of variability were, the sentence - 26.3%, the talker - 4.9% and an interaction of talker and sentence – 5.1%, the remaining variance was accounted for by residual error – 51.2%; Demorest & Bernstein, 1992). This estimate was generated using generalisability analysis, which essentially partitions recorded speechreading scores into separate compartments in order to identify the causes of variability. In an effort to remove such variability and improve speechreading performance as a whole, several training studies have been developed. Such studies are of great importance for their potential to improve the speechreading performance of deaf observers, thus enhancing their ability to communicate with normal-hearing people.

One of the first visual speech training studies investigated the recognition of consonants (Walden, Prosek, Montgomery, Scherr & Jones, 1977). Each consonant was paired with the vowel /a/ in order to produce consonant-vowel (CV) pairs. The training sessions were 14 hours in length, split into 14 one-hour blocks and spread over two weeks. In each session, participants were asked to perform a same / different discrimination task or to discriminate between CV pairs. Every participant was tested on their ability to discriminate visual consonants before and after the training sessions. The results were plotted based on the viseme groupings described earlier. A correct response was one which was included within a particular viseme cluster as opposed to a correct phonetic response. Training was found to increase the number of visemes used by participants and improve viseme discrimination by 25% (from 68 – 93% correct). The greatest increase in performance occurred within the first hours of training, with any further improvement becoming progressively smaller in size.

A second visual speech training study examined the effects of visual consonant discrimination training on the intelligibility of audiovisual sentences for hearing-impaired participants (Walden, Erdman, Montgomery, Schwartz & Prosek, 1981). This study used a similar training technique to Walden et al (1977) but over a shorter time span (7 hours of training in 14, thirty-minute sessions). The authors found a mean improvement in viseme discrimination of 10% (83% to 93% correct). The same group of subjects was consequently found to have a 26% (from 38.5% to 64.5% correct) improvement in audiovisual speech-in-noise intelligibility. The authors suggest that the difference in improvement between the two studies (25% versus 10% increase in accuracy) was due to talker variability, with the talker in the second study being more difficult to speechread than the talker in the first study. However, overall both studies indicated that training can improve visual speech discrimination of small units of speech, measured through CV pairs.

A later study looked a speechreading training using CV syllables once more, but this time participants were also tested on their ability to discriminate visual words (sets of ten visemic minimal pairs that differed by only one articulatory movement e.g. /pin/ and /tin/) (Gesi, Massaro & Cohen, 1992). The aim was to examine if training would transfer from one task to another. The participants were split into two groups. One group received detailed and explicit instructions regarding the task and information about visible speech signals. The second group were given only basic instructions about the task. The results indicated no significant difference between the groups' performance and both showed a significant improvement in visual CV discrimination. As before, the majority of this learning effect occurred within the first two blocks of trials, or the first hour of training. The results of the transfer test showed that trained participants were not significantly more accurate than participants in a control group who had received no training. Thus, it would appear that training can improve CV discrimination, but the results of that training do not transfer to the visual discrimination of words. The authors suggest this lack of transfer could be due to differences in testing method between the training and transfer phases, or due to a change in talker between the two phases (Gesi et al, 1992). Alternatively, it is possible that an increased ability to discern small units of speech does not improve lexical speechreading ability, but further research is required to test this hypothesis.

Two further studies by Bernstein and colleagues (Bernstein, Auer & Tucker, 1998; Bernstein, Auer & Tucker, 2001) investigated the use of vibrotactile aids to facilitate visual speech training and the use of short-term training to enhance speechreading. The first of the studies (Bernstein et al, 1998) trained both hearing-impaired and normal-hearing subjects using sentences, with performance measured using keyword identification accuracy. Approximately seven hours of training was given to each participant, spread over three weeks. Results showed an average increase of 10% accuracy in the normal-hearing subjects and only a 5% increase in the hearing-impaired subjects. Similar to an earlier experiment (Walden et al, 1977) most of the training improvement occurred in the initial stages. No advantage was found for using a vibrotactile aid.

Their second study (Bernstein et al, 2001) investigated the effects of using short-term training (with sentences) in an attempt to improve the speechreading abilities of normal-hearing subjects up to the level of the hearing-impaired, as the hearing impaired are generally found to be more accurate at speechreading. Training involved presentation of a visual sentence, followed by the correct text version of the sentence on screen as feedback. The test sessions presented the visual stimuli without feedback. Their results indicated only a small, but significant, increase with training of between 5 – 7% for both groups. Again no advantage was found from using the vibrotactile aid. This small increase in performance did not significantly reduce the difference in performance between the normal-hearing and hearing-impaired participants as illustrated in Figure 1.10. Thus, short-term training was not sufficient to improve the performance of normal-hearing participants to a level on a par with the hearing-impaired participants.

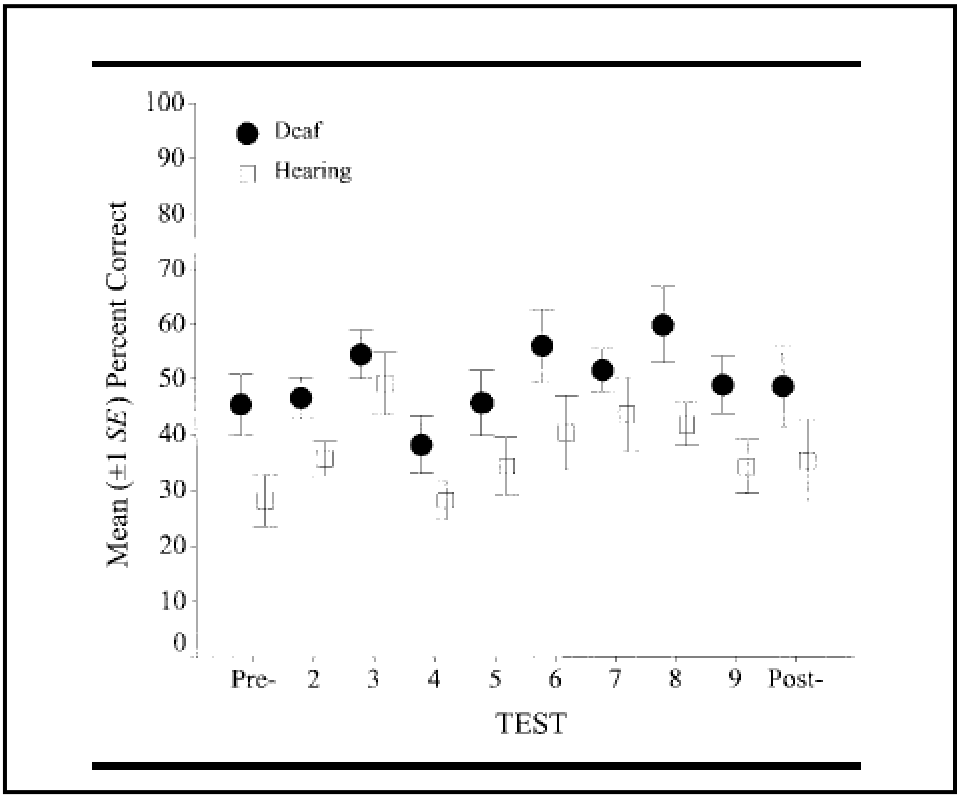


Figure 1.10: Percent words correct in terms of group (hearing-impaired and normal-hearing) and test (pre, training and post test), shown with standard error bars (Bernstein et al, 2001).

In summary, the majority of speechreading training experiments thus far show a relatively small, but usually significant, increase in speechreading performance. The smallest amount of improvement is associated with training using sentence materials, suggesting that it may be easier to train an individual to distinguish between simple visual stimuli, such as CV syllables, than more complex linguistic stimuli such as sentences. However, the lack of generalisation between CV syllables and sentences means that sentences remain a viable training material.

1.9 Overview of the main research questions

A review of the literature has highlighted the many similarities between auditory and visual speech processing. Auditory and visual intelligibility are both adversely affected by talker variability and conversely benefit from talker familiarity. The two modalities complement each other in terms of speech perception, with audiovisual speech perception being associated with fewer errors than either modality alone. This integration of the two sets of speech information has led to my formation of hypotheses regarding accent variation and visual speech, based upon the heavily documented effects of accent upon auditory speech.

The main prediction tested in this thesis is that accent type will influence the appearance of the visual signal and therefore potentially visual speech intelligibility. This prediction is supported by previous research which indicates a link between face processing and speechreading (Yakel et al, 2000). That research suggested that talker-specific information is encoded alongside phonetic information relating to speech production. Since accent type could potentially influence both facial movements and the corresponding distinctiveness of visemes, it is likely that accent will prove to be an important talker characteristic. A more specific prediction is that an unfamiliar accent type will have an adverse effect upon speechreading performance, in much the same way as an unfamiliar talker impairs auditory speech intelligibility.

Chapter 2 reports an exploratory study on accent discrimination utilising foreign accent (French). Further experiments in Chapters 2 and 3 determine whether there is any effect of non-native *or* regional accent type upon speechreading performance, either due to familiarity with the accent or characteristics associated with the effect of accent upon visible articulation. If the visual speech system operates in the same way as the auditory system we can expect to see a drop in performance associated with an unfamiliar accent. Chapter 4 focuses upon the combined influence of talker variability and accent variation on speechreading performance. Essentially, the cognitive effort of encoding accent characteristics was expected to produce an additive effect when combined with talker variation, exacerbating the effect of multiple-talkers on speechreading performance. Finally, Chapter 5 focuses upon the mediation of accent effects through exposure, contextual cues and repetition. Auditory research indicates that a relatively short period of exposure is sufficient to compensate for the effect of an unfamiliar accent upon speech perception. The aim of Chapter 5 was to determine the requirements for successful mediation of accent effects within the visual speech processing system.

Chapter 2. The Effect of Foreign Accent on Discrimination and Intelligibility of Visual Speech

'The accent of one's birthplace remains in the mind and in the heart as in one's speech'
Francois de la Rochefoucauld, 1680.

The goal of many second language (L2) learners is to produce speech in the manner of a native talker (Munro & Derwing, 1995). However, the majority of L2 learners usually produce their L2 with a variety of pronunciation (e.g. alteration of vowel or consonant sounds, incorrect use of stress), grammatical (e.g. incorrect use of plurals, inappropriate verb tense; Munro & Derwing, 1995) and vocabulary errors (e.g. inappropriate word usage), the extent of which will depend on the learner's level of fluency (Gass & Varonis, 1984). These errors relate to the non-native speaker utilising features of their first language (L1), in their production of L2 (Flege, 1984), essentially mapping constraints and phonetic rules of L1 onto the production of L2. Previous auditory research indicates that grammar and pronunciation errors produce the most difficulties in speech intelligibility (Gass & Varonis, 1984). Thus, Chapter 2 focuses on the effect of non-native pronunciation (foreign accent) upon speech intelligibility.

Non-native or foreign, spoken accent relates to a departure from the phonetic norms of a language through the speaker utilising unfamiliar (to a native speaker) realisations of speech sounds (Flege, 1984). These realisations occur on both the segmental (phonemic errors) and supra-segmental (intonation, rhythm and lexical stress) level (Neri, Cucchiarini & Strik, 2006). There is little research on the specific impact of these errors upon non-native speech production, but the general view is that the resulting irregularities in the speech signal could impede effective processing and reduce auditory speech intelligibility (Clarke & Garrett, 2004). Certainly auditory research has shown non-native speech to have a detrimental effect upon transcription of speech sounds (Gass & Varonis, 1984) and the judgement of pronunciation errors (Schmid & Yeni-Komshian, 1999), both indicators of speech intelligibility.

Foreign accent has also been shown to adversely affect both sentence comprehension (Munro & Derwing, 1995) and processing speed (Clarke & Garrett, 2004). The first of these studies asked English-speaking participants to assess the truth level of various true or false statements, each uttered in English by a talker with a Chinese-Mandarin or English accent (Munro & Derwing, 1995). Statements spoken in a Mandarin accent had a detrimental effect on decision processing speed. In the second study, Clarke and Garrett (2004) examined the reaction times of native English speakers to non-native renditions of English (Spanish speakers producing English sentences). Performance was measured by recording a participant's responses to visual probes (text words appearing on screen) the task being to identify the visual probe that matched the last word of a preceding spoken sentence. The sentences were presented in blocks, four sentences per block. There were three experimental groups, i) accent; participants were presented with auditory sentences spoken in English by a Spanish speaker, ii) control; participants were presented with 12 sentences spoken by a native English speaker, followed by four sentences spoken by the Spanish speaker, iii) no-accent; participants were presented with sentences spoken by a native English speaker. The results, as illustrated in Figure 2.1, showed that initially the matches in the no-accent and control groups were significantly faster than the accent group.

However, after each group had been exposed to 12 sentences, the accent group showed a significant improvement in performance with reaction times recorded at a similar speed to those for the control and no-accent groups. Finally, the control group were significantly slower at responding to the final four sentences of the experiment than the no-accent and accent groups.

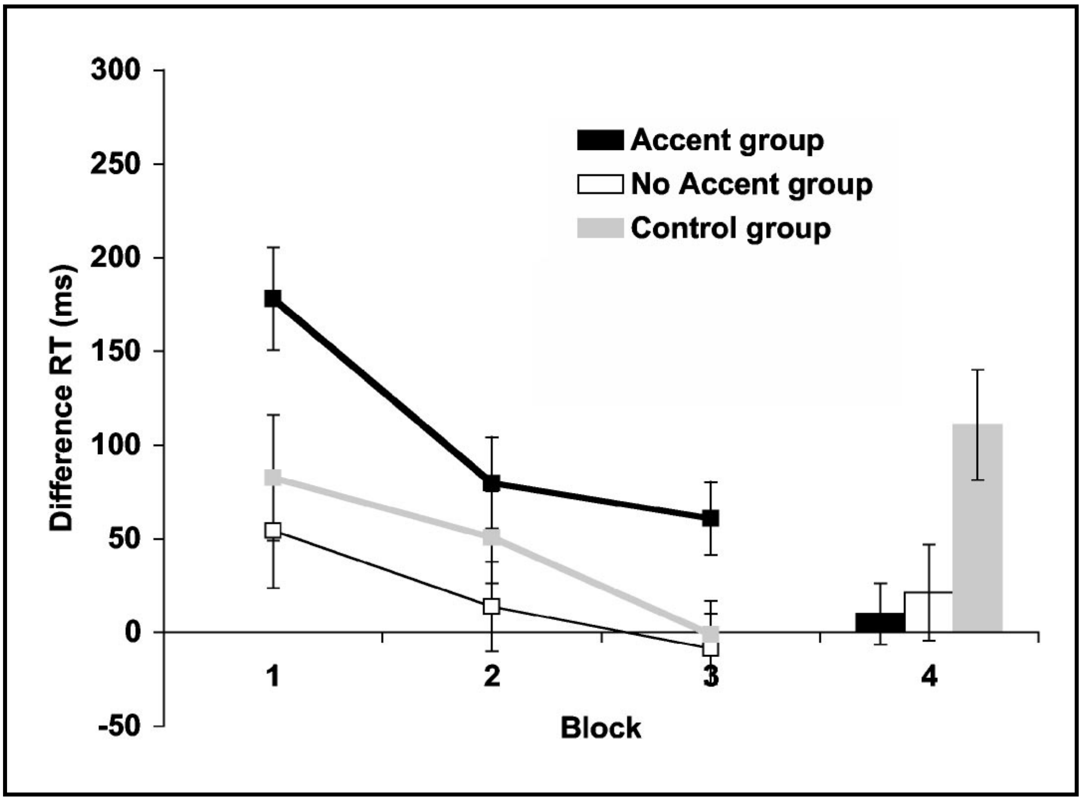


Figure 2.1: Mean difference reaction times according to group, each block consists of 4 sentences (Clarke & Garrett, 2004)

The recorded increase in reaction times for the control group in block four indicates that the improvement seen in the accent group across all four blocks incorporated adaptation to an unfamiliar accent and were not simply due to practice effects. The study illustrates the adaptability of the auditory system, with learning and consequential alteration of phonetic categories occurring within one minute of exposure (12 sentences) to non-native accented speech (Clarke & Garrett, 2004).

There has been no previous research on the effect of foreign accent upon visual speech intelligibility. However, research into the utilisation of visual cues in L2 learning (Davis & Kim, 1998; 2001) and the discrimination of language through visual speech (Soto-Faraco et al, 2007) offer some insight as to the potential effects of non-native speech on visual speech intelligibility (see Chapter 1, Section 1.5). Certainly it would appear that observers are able to utilise the visual signal to improve intelligibility of speech produced by non-native talkers or to discriminate language types. This suggests that the visual signal provides meaningful information about language and potentially also accent type. A consequence of this is that the contingent alterations to the appearance of the visual signal may have an effect on an observer's ability to speechread that talker. The research reviewed thus far would appear to predict that foreign accent could potentially impair speechreading ability.

To summarise, previous research indicates that foreign accent alters auditory speech through unfamiliar phonetic realisations and the expectation is that such alterations will also influence speechreading performance via the visual signal. Chapter 2 provides a first step towards quantifying the effect of accent upon speechreading performance. Thus, the purpose of this chapter is to investigate the discrimination of accent through visual speech and the effects of foreign accent upon speechreading performance. The hypotheses are as follows:

- Discrimination of language and accent should be possible using the visual modality alone.

- Foreign accent should have a detrimental effect on visual speech intelligibility due to the inclusion of unfamiliar articulatory motions associated with L1 constraints

2.1 Experiment 1: The discrimination of language and accent through visual speech

Experiment 1 provides a first step towards quantifying the effect of accent upon speechreading performance by investigating the ability of individuals to discriminate accent using visual speech. Comparable research in the auditory domain has already shown that listeners are capable of discriminating accent from sentences, phrases and syllables, with a 63 – 95% accuracy rate (Flege, 1984). Listeners are extremely sensitive to any deviation from the phonetic norms of their native language, allowing them to discriminate foreign accent quickly and accurately (Flege, 1984). The ability of observers to utilise the visual signal in accent discrimination has not been previously examined. However, research which examined the perception of non-native phonemic contrasts using auditory and audiovisual stimuli found a significant advantage in the perception of a labial (/p/) versus a labiodental (/v/) place of articulation when the stimulus was audiovisual (Hazan, Sennema, Faulkener, Ortega-Llebaria, Iba & Chung, 2006). The authors suggest that when a contrast between speech sounds is visually salient the visual cues will be utilised by a L2 learner to aid speech perception (Hazan et al, 2006). This indicates that visual cues contain supplementary information relating to language norms which can be used by an observer to improve his/her ability to distinguish speech sounds. It is possible that those same visual cues may provide information relating to language and accent, allowing observers to discriminate between accents visually.

When designing Experiment 1, an important consideration was that the visual intelligibility of speech varies considerably between talkers, even when accent is not a factor (Kricos & Lesner, 1982, 1985). This could potentially make it difficult to isolate accent from the influence of any other co-varying factors between talkers, such as idiosyncratic variation in speech rate, mouth opening etc. The characteristics of the talker must be kept constant, therefore, in order to ensure that any judgements are based upon accent and not talker variation (Magen, 1998). Soto-Faraco et al. (2007) solved this problem by having the same talker produce speech in two languages (Spanish and Catalan) in their recent paper. The same approach is used in Experiment 1, in which the same bilingual talker produced two languages and accents in a 2 x 2 factorial design. It was important to use a bilingual talker to ensure complete fluency in the production of both accents and languages. It should be noted at this point that the aim here was not to test whether the French language was more difficult to speechread than the English language. Indeed native French observers are able to speechread French at least as well as native English observers can speechread English (Gagne, Charest, Le Monday & Desbiens 2006). It is the effect of an unfamiliar accent upon speechreading performance of English speakers that is the focus of Chapter 2.

French and English were chosen as comparison accents due to well-established rhythmic and pronunciation differences between them. These differences take several forms. First, in terms of speech prosody, English is a stress-based language, which means that the rhythm of a talker's speech is based on stressed syllables appearing at a roughly constant rate with non-stressed syllables being shortened (Carr, 1999). French, in comparison, is a syllable-timed language; that is, syllables appear at a constant rate regardless of stress. Lexical stress is associated with greater jaw opening displacement, greater lip closing displacement and longer jaw opening durations (Tye-Murray & Folkins, 1990). To an observer who is an English native speaker, the pattern of speech movements associated with a French accent should, therefore, clearly differ from those associated with a native English accent in terms of perceived visual word stress and associated segmentation.

Second, in terms of articulation, the two languages differ in both vowel and consonant pronunciation. For example, many English speakers struggle to comprehend and reproduce the French front rounded vowel 'y' as it is not something they produce in their own language. There also exists in English the phenomenon of 'vowel reduction' (Tranel, 1987), whereby an unstressed vowel loses its prominence within a word. Such a reduction is not present in French and so unstressed vowels are more prominent in French than in English (Tranel, 1987). The two languages also differ in the place of articulation of certain consonants. For example, in the French and English production of 's', the French pronunciation is dental, and the English alveolar, with pronunciation originating from further back in the mouth (Flege & Hillenbrand, 1984). Thus, the visual appearance of English words should be altered when produced in a French accent, potentially reducing speechreading accuracy.

Experiment 1 determined whether native English speakers can differentiate between their own familiar English accent and language, and the less familiar and foreign French language and accent, using only the visual modality. This was an important first step in our investigation because our suggestion that accent influences visual speech rests upon the assumption that accent variation alters the form of visual speech. If there are no distinguishable differences between the English and French accents then the claim that accent will have an impact upon speechreading performance would be weakened. In contrast, if the change in accent type is discriminable then this upholds our proposal that accent change is associated with a visible alteration in the appearance of visual speech.

2.1.2 Method

Participants

Thirty adult participants aged between 18 and 35 (mean age: 22) took part in Experiment 1. All were English native speakers and had been born in England. All participants reported normal or corrected to normal vision and a normal level of hearing. All participants reported a GCSE level of experience with the French language (i.e. five years of secondary school education in French). The talker used was not familiar to any of the participants.

Stimuli

Fifty sentences from the BKB (Bench, Kowal & Bamford, 1979) sentence materials were recorded. The BKB materials consist of a set of short sentences such as *'the puppy played with a ball'*. They were constructed using simple vocabulary and have been used in many tests of speechreading ability (see MacLeod & Summerfield, 1987 for a summary). Recordings were made in each of four conditions: English spoken with an English accent, English spoken with a French accent, French spoken with an English accent and French spoken with a French accent, giving a full factorial design. The sentences used in each condition were exactly the same, with the French stimuli being direct translations of the English materials. The talker was a bilingual, 24 year old, English male who was born in England and had spent eight years (from ages 10 to 18) living in the Alsace region of France. The talker returns to France at least once a year and retains his knowledge of French through frequent use of that language with French colleagues at work, talking to French friends on the phone and watching French films.

In the recordings, the talker's face was fully illuminated using three high-power lamps placed at right angles to reduce shadowing. Recordings were made against a neutral white background with only the face and neck of the talker visible. Each recording featured his full face filmed from a camera (Sony Digital Camcorder, DSR-200AP) placed 1.5 meters away and directly in front of him. Each recorded sentence was preceded by 1s of the talker's static face in a closed-mouth position, and followed by a further 1s of static footage again in the closed-mouth position. The auditory signal was not recorded.

Procedure:

Each condition contained 50 visual sentences and was randomly presented giving 200 clips in total. The process of randomisation included the proviso that no condition was presented twice in a row, ensuring trial-by-trial variation. Participants were split into two groups, each given a different task. Group 1, the 'language' group were asked to discriminate the language, Group 2, the 'accent' group were asked to discriminate the accent type. The groups differed only in the instructions that they were given, the stimulus set was the same in both cases.

Each participant was seated at a table directly in front of the screen at an approximate distance of 70 cm and instructed to watch each video clip carefully. Participants were not explicitly required to comprehend the individual words, only to make a language or accent discrimination. Responses were made using a two key keypad; Key 1 for 'English', Key 2 for 'French'. There were four practice trials before the experiment began, in which one example from each condition was given.

2.1.3 Results

Figure 2.2 shows discrimination performance (number correct from possible total of 200) for both 'language and 'accent' groups. Performance levels in the 'accent' group were lower (mean: 58% = 115 keywords correct) than for the 'language' group (mean: 63% = 127 keywords correct) though this difference failed to reach significance ($t(28) = 1.837$, $p > 0.05$). Performance in both groups was at a level significantly greater than chance; 'language' group: ($t(14) = 4.556$, $p < 0.05$), 'accent' group: ($t(14) = 7.776$, $p < 0.05$).

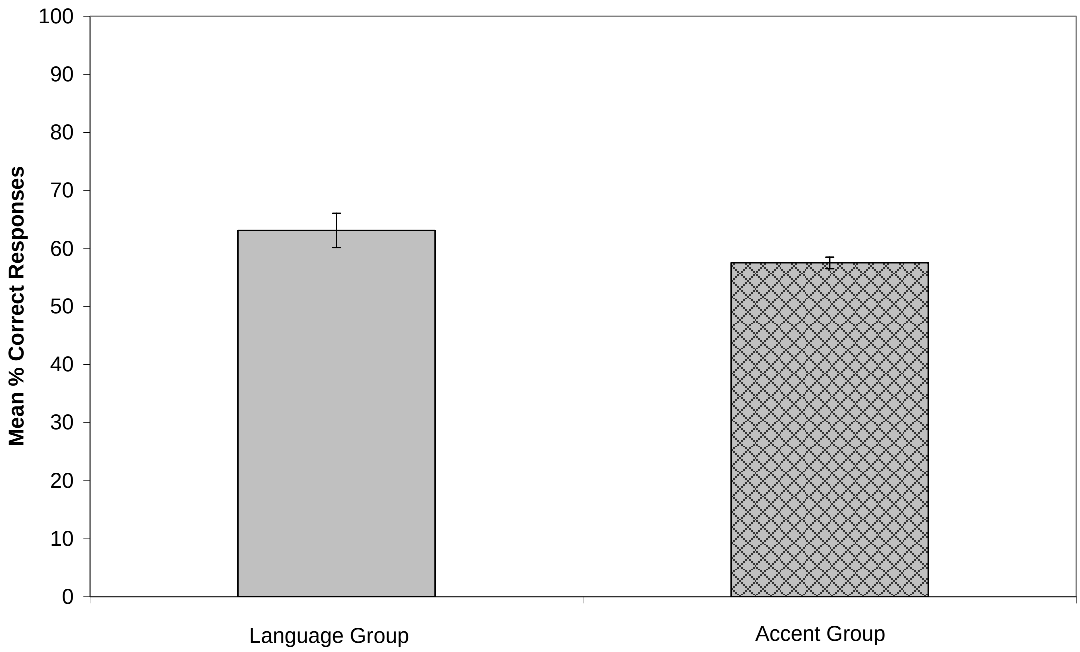


Figure 2.2: Mean percentage of correct discriminations of French or English language or accent type, shown with standard error bars

When performance levels of both groups was reanalysed according to condition, differences in performance became apparent. These were related to the congruence (language and accent match) or incongruence (language and accent mismatch) of the stimulus. By collapsing across conditions, a score was produced for each group when language and accent were congruent (i.e. English language, English accent and French language, French accent) and when they were incongruent (English language, French accent and French language, English accent). When the stimuli were congruent, the task could be successfully completed using either language or accent cues, irrespective of the instruction given. In comparison, when the stimuli were incongruent, the observer had to extract the relevant factor (language or accent) from the signal, and base his/her discrimination upon that factor alone. A drop in performance associated with incongruent stimuli would indicate that an accent-language mismatch alters the visual signal and in turn has a detrimental effect on language or accent discrimination.

A two-factor mixed factorial (2 group x 2 congruence) ANOVA revealed a main effect of congruence (Figure 2.2): $F (1, 28) = 79.01$; $p < 0.01$. Performance for the incongruent stimuli (mean: 52% correct) was significantly lower than those for the congruent stimuli (mean: 70% correct). There was no significant effect of group, indicating no overall difference in performance between the two discrimination tasks. However, there was a significant interaction between group and congruence: $F (1, 28) = 14.333$; $p < 0.01$.

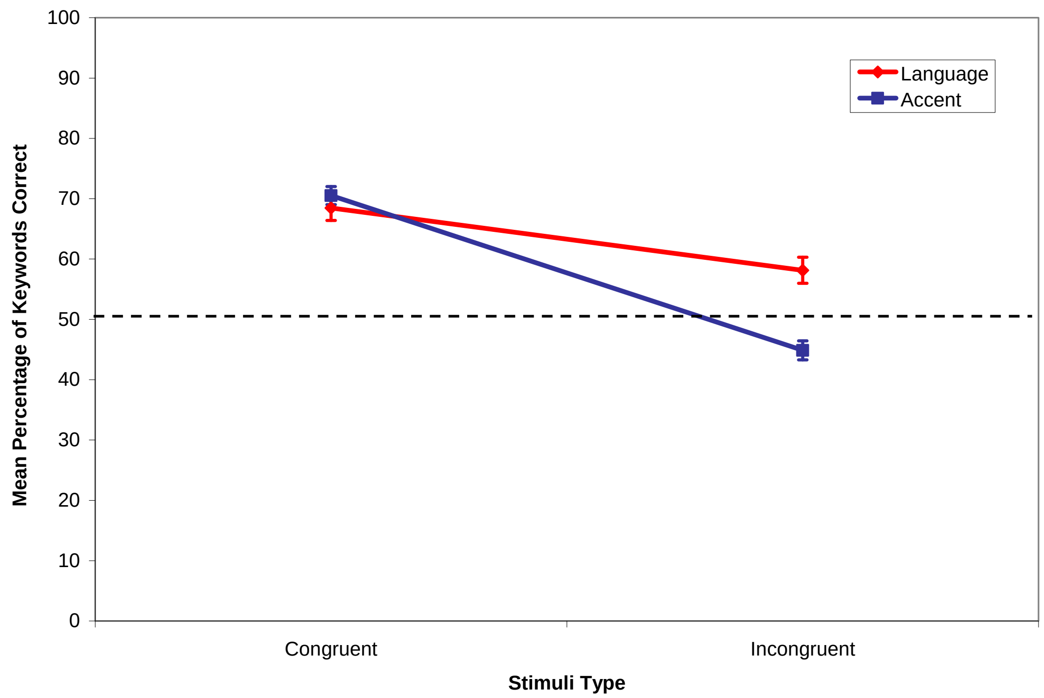


Figure 2.3: Mean percentage of correct discriminations of language or accent in congruent and incongruent stimuli, shown with standard error bars

The interaction indicates that the mismatch between language and accent exerted a more pronounced effect on the 'accent' group. Indeed, Figure 2.3 shows that the performance level for the accent group is below chance (45%) for the incongruous stimuli.

2.1.4 Discussion

Observers were able to discriminate between French and English based on a unimodal visual signal. This result supports the hypothesis that a talker's visual articulation is affected by accent and language, and that the differences are visually discriminable. This interpretation is further strengthened by the use of the same talker in all conditions, ensuring that the only possible cue for language and accent discrimination is the visual signal. Therefore, participants were able to use the changes in visible articulatory movement to distinguish between the languages and accents.

The results are consistent with the recent work on language discrimination, using Spanish and Catalan, reported by Soto-Faraco and collegues (2007). The authors argued that participants used a combination of cues (phonetic, rhythmical and lexical differences) in order to discriminate between the two languages. It is likely that similar cues were used by the participants in Experiment 1. As discussed earlier, French and English vary along several dimensions including stress patterns, pronunciation, rhythmic structure and vowel length, all of which could be utilised as discriminatory cues. Moreover, French and English differ along more parameters than Spanish and Catalan making the task of discriminating between them potentially easier (Soto-Faraco et al, 20007). This suggestion is upheld by our results for the congruent stimuli (68% correct for language group) where performance was higher than that recorded by Soto-Faraco and colleagues (2007) (57-61% correct).

Our reanalysis of the data revealed an interesting cost for the incongruent stimuli. For the group discriminating accent, an accent / language mismatch severely impaired performance (45% correct for incongruent stimuli, 71% for congruent stimuli). This condition required participants to select the relevant features of accent independent from those of language discrimination. The recorded effect indicates that the participants were unable to do this accurately, resulting in a drop in performance. Alternatively, it is possible that the participants were basing their decisions about accent type on the language which they deduced the talker to be producing, giving more weight to actual vocabulary (i.e. assuming that if the talker was producing the English language he would be using an English accent) as opposed to the phonetic (pronunciation) and rhythmic cues.

For the group discriminating language, an accent / language mismatch also impaired performance (58% correct for incongruent stimuli, 68% correct for congruent stimuli). This condition required the participants to select the relevant features of language independent from those of accent. Although accent itself was difficult to *explicitly* determine, it did *implicitly* affect language discrimination performance. Thus, an incongruent accent seems to alter certain key visual cues on which language judgments are based. This result is consistent with previous findings that visual cues offer supplementary information when attempting to discern L2 phonemes (Hazan et al, 2006). Phoneme discrimination requires perception of the place of articulation and thus incorporates accent type. A mismatch between language and accent could provide misleading information about place of articulation, generated by the mapping of L1 phonetic structure onto the L2, potentially reducing performance.

In summary, observers were able to visually distinguish two languages and accents produced by the same talker. When the language and accent did not match performance was compromised. The interpretation of this result is that some of the features on which language and accent discriminations are based are an integral part of the visual speech signal and the cues for language and accent cannot be separated. If regional accents were used, removing any language differences, accent discrimination could be isolated. This issue is examined in Experiment 5.

2.2: Experiment 2: The effect of accent variation (French / English) upon speechreading ability

The results from Experiment 1 indicate that changes in articulatory patterns associated with differing languages and accents are distinguishable, though language discrimination is more successful than accent discrimination. Experiment 2 investigated whether the differences in articulatory patterns associated with accent influence speechreading ability. Essentially, we aim to determine whether the same detrimental effects associated with non-native speech productions within the auditory domain (Munro & Derwing, 1995) are replicated here in the visual domain.

In order to investigate the effect of foreign accent upon visual speech, a comparison was made of speechreading performance using English language sentences spoken in either an English or French accent. The specific hypothesis was that speech produced in a French accent would be less intelligible than speech produced using an English accent for native English speakers.

This hypothesis was based on earlier visual speech research on talker variability. It has been shown that viseme production, and hence visual speech readability varies from talker to talker (Kricos & Lesner, 1985). It is well established that the production of L2 by a non-native talker will alter the phonetic realisation of that language due to mispronunciations relating to the talker's L1 knowledge (Flege, 1984). These alterations should also influence that talker's viseme categories, and thus their intelligibility, potentially having a detrimental effect on speechreading performance. It has also been shown that the rhythmic differences between French and English, some of which relate to lexical stress make it more difficult for native English speakers to successfully distinguish visual word boundaries in accented speech. This may also have a detrimental effect on visual speech intelligibility.

There are two final considerations, first, it should be noted that the auditory signal carries information relating to accent that cannot be conveyed by the visual signal. Essentially, voicing differences and aspects of accented speech which are produced using the articulators at the back of the mouth, will not be visible (Hazan et al, 2006). Thus, it is possible that the detrimental effect of foreign accent upon visual speech will not be of the same magnitude as that upon auditory speech. Second, listeners are able to swiftly adapt to an unfamiliar accent and rapidly improve their performance (Clarke & Garrett, 2004). This typically occurred within one minute of exposure to the unfamiliar accent type. Thus, it is possible that participants within the present experiment may become able to encode the articulatory gestures of the French accent during the test period. Both of these factors could lead to a null effect of foreign accent on speechreading performance.

2.2.1 Method

Participants

Twenty adults participated in Experiment 2, all native English speakers, none of whom had participated in Experiment 1. All participants reported normal or corrected to normal vision and good hearing. All participants reported GCSE experience in French (five years of secondary school education in French). The talker was not familiar to any of the participants.

Stimuli

A total of 140 BKB sentences were spoken by the same talker as in Experiment 1. These formed three lists, i) 40 sentences all spoken in an English accent that formed a basic speechreading measure, ii) 50 sentences spoken in either a French or English accent and iii) a further 50 sentences spoken in a French or English accent. Lists ii) and iii) were the experimental set and were fully counterbalanced so that half of the participants viewed list ii) spoken with an English accent and list iii) spoken with a French accent and vice versa.

Procedure

Each participant was seated at a table directly in front of the computer screen. They were instructed that the talker would produce one sentence per video clip, which they were asked to watch carefully. The task was to report the sentence that the talker had said, typing their response into a computer keyboard. They were not required to report the entire sentence; any word that was typed in was recorded. Experiment 2 consisted of three sections. First, each participant completed a short practice session that contained two video clips. Second, they completed the speechreading measure, consisting of 40 video clips presented in random order. Third, each participant completed the speechreading experiment with 100 sentences (lists ii and iii) presented in random order so that trials varied from one accent to the other in a fully randomised manner.

2.2.3 Results

Scoring used a loose keyword scoring procedure where errors in morphology are ignored (Bench, Kowal & Bamford, 1979). Each sentence had three keywords, and a point was awarded for each correctly identified keyword. A participant's score therefore represents the percentage of correctly identified keywords within a sentence list. Each score in Figure 2.4 represents the percentage of keywords correctly identified out of a potential total of 300.

The majority of participants were more accurate when speechreading sentences spoken with an English accent (mean: 26% correct) than for those spoken in a French accent (mean: 19% correct). A paired t-test showed the difference to be significant ($t(19) = 4.902$, $p < 0.005$).

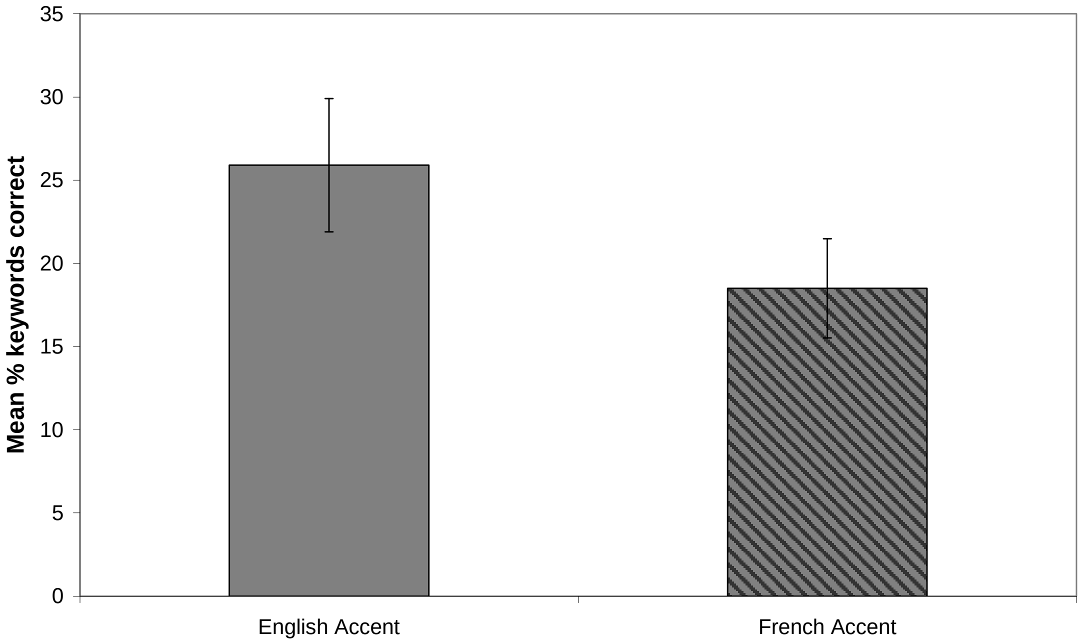


Figure 2.4: Mean percentage of correctly identified keywords from sentences produced in either a French or English accent, shown with standard error bars

A mean split was performed on the scores of the basic speechreading measure to separate the groups into "poor" and "good" speechreaders. Mean performance was 21% and ten participants had scores at or above 21% and so were designated "good" speechreaders, the remaining ten participants (< 21%) were designated "poor" speechreaders. Figure 2.5 illustrates the mean percentage of correctly identified keywords by each group.

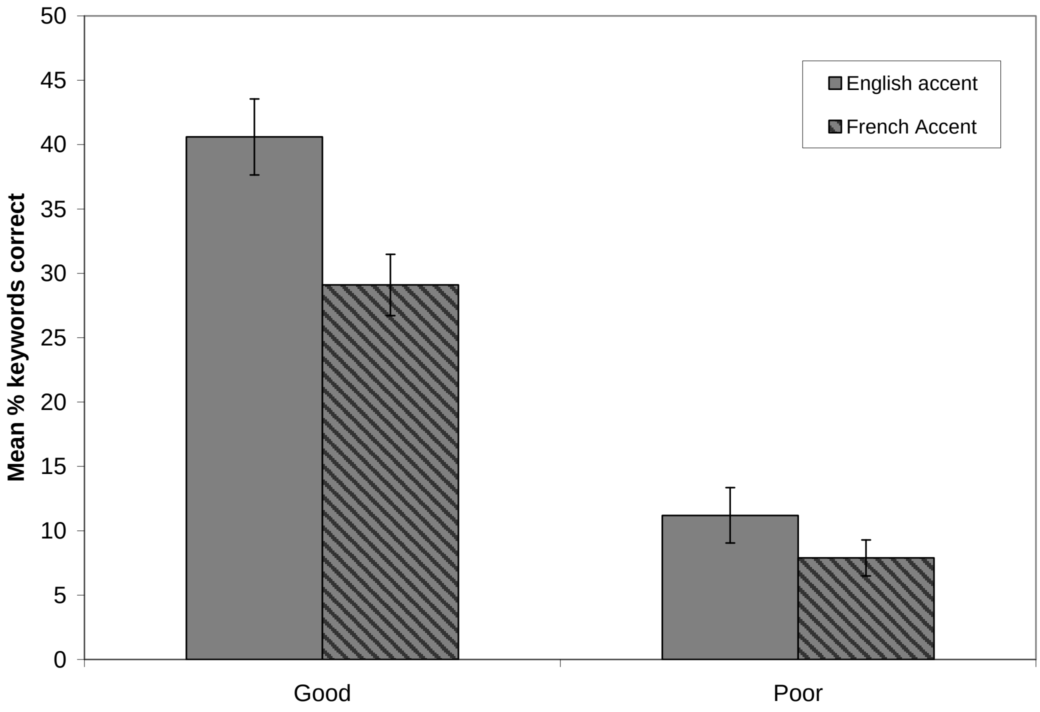


Figure 2.5: Mean percentage of correctly identified keywords by the designated "good" and "poor" speechreaders, shown with standard error bars

A two-factor mixed factorial (2 speechreading ability x 2 accent) ANOVA, showed that the performance values for the good and poor speechreaders differed significantly: $F(1, 18)$ 129.034, $p < 0.01$. There was also a significant interaction between accent and speechreading ability: $F(1, 18)$ 11.355, $p < 0.01$. Thus, there was a greater effect of accent upon the performance of the good compared to the poor speechreaders. Post-hoc analysis confirmed this by indicating that the difference in keyword identification accuracy shown by the good speechreaders was significant ($p < 0.05$), whilst the observed difference in performance shown by the poor speechreaders was not ($p > 0.05$). The lack of an effect for the poor speechreaders may be influenced by floor effects.

2.2.4 Discussion

The results showed a detrimental effect of the French accent upon visual speech intelligibility for English language sentences. This indicates that the unfamiliar visible articulations produced by mapping phonetic features of the French language onto English speech were sufficient to reduce the intelligibility of the visual speech signal. These results in the visual modality are similar to those found in the auditory modality (Munro & Derwing, 1995). One methodological advantage of the present experimental design is the use of a single talker for both accent types. Hence, no other factor relating to talker variability can be responsible for the observed results other than accent. There are two potential explanations for the effect of foreign accent on speechreading ability; the encoding of talker-specific information and the formation of visual speech prototypes.

'Encoding' theory (Mullenix, Pisoni & Martin, 1989), as described in Chapter 1, Section 1.3, argues that an individual's perception of speech adjusts each time he/she comes into contact with a new pattern of speech sounds, such as a new talker, or an unfamiliar accent. This adjustment requires cognitive resources and so may impair perceptual processing. In Experiment 2, a cognitive cost would have been caused by the unfamiliar aspects of the visual speech signal produced by the talker using a foreign (French) accent. The increased cognitive demands involved in encoding this talker-specific variation could lead to the observed decrement in performance.

An alternative explanation from the field of auditory speech perception concerns the formation of speech prototypes, or, in terms of accent, the formation of expectations about how speech should be ideally pronounced to facilitate comprehension (Flege, 1992). Individuals form a 'tolerance region' around these prototypes as their experience of speech increases (Flege, 1992), this allows them to identify differences in sound production as ‘foreign’ and then process those sounds in a meaningful way. The participants in Experiment 2 were not informed that the talker would be varying his accent. It is possible, therefore, that although they each had experience of the French language, they did not categorise the variation in the visual signal as accent induced, as this was outside their expectations of the task. As such, they may have been unable to process the differences in a meaningful way resulting in the recorded drop in performance.

It should be noted at this point that the basic measure of speechreading (list i) was presented in an English accent. It could be argued that this initial exposure to the English accent produced a practice effect that led to the better performance for this accent. However, although there is a small increase in performance (from 21% to 26%) across lists, this difference is smaller than the effect of accent manipulation. The improvement could equally be explained by the increase in material shown to the observer, from 40 sentences in the speechreading measure, to 100 sentences in the main experiment. Either way the gain in performance was too small to be the only cause for the observed detriment in performance for those sentences pronounced with a French accent.

The results indicate that those participants designated good speechreaders were more affected by the French accent than the poor speechreaders. Therefore, it would appear that better speechreading ability does not necessarily allow the observer to fully adapt to unfamiliar accents over the course of the testing period. Moreover, despite the recorded effect size, the performance of the good speechreaders for the sentences produced in a French accent was still significantly more accurate than that of the poor speechreaders. Thus, although an unfamiliar accent impaired speechreading performance it did not reduce it to the level of the poor speechreaders. It is possible, therefore, that a good speechreader may be able to compensate for some of the effects of accent when communicating with unfamiliar talkers in an everyday environment. This suggestion is examined in more detail in Chapter 5, where various methods are utilised in an attempt to mediate accent effects.

Although it was considered advantageous to use the same talker to utter the sentences used in Experiment 2, arguments can be made against this methodology. For example, the French accent produced could be considered as just a distortion in visible articulation rather than as an authentic accent. The use of a single talker may also reduce the ecological validity of the results. However, there is evidence to suggest that intra-talker variability in accent production is an important consideration. Accent has been shown to change according to their exposure to different accent types through regional relocation and through conscious alteration in order to raise their social standing by, for example, adopting an HC (Home Counties) English accent (Howell, Barry & Vinson, 2006). Thus, although using a single talker perhaps reduces our ability to generalise the results, they are not without merit when considered in the light of intra-talker variability. Finally, having determined that a produced accent change effects the ability of observers to speechread one talker, further experiments using multiple talkers of different accents will be conducted to strengthen the argument that accent has an impact upon visual speech.

2.3 General Summary

The results of Experiment 1 indicate that accent is a salient aspect of the visual speech signal. Certainly participants were able to discriminate between languages and accents using the visual modality at a level above chance. Thus, it would appear that the visual signal is significantly altered by a talker's language or accent type, producing visual cues that an observer is able to decode in order to distinguish between different language and accent types.

Experiment 2 shows that intra-talker variations in accent type can have a direct influence on visual speech intelligibility. Specifically, the rhythmical and pronunciation changes produced by a talker speaking in a French accent were sufficient to have a detrimental effect on an observer's ability to speechread that talker.

These results represent the first quantitative examination of accent effects in the visual modality and are indicative of the potential importance of accent as a visual talker characteristic.

Chapter 3. The Effect of Regional Accent on the Discrimination and Intelligibility of Visual Speech

'You like tomato, and I like tomahto…you say laughter, and I say lawfter'

(Ira Gershwin, Let's Call the Whole Thing Off)

Speech characteristics associated with both native language and regional accent exert a significant effect on auditory and visual speech (Wells, 1982a; 1982b). Chapter 2 investigated the effect of foreign accent on the visual discrimination of accent type and speechreading performance. Chapter 3 investigates regional accent effects upon similar visual tasks. Previous research in the field of auditory speech indicates that accent variation occurs along a continuum with foreign accent at one extreme and regional accent variation found at the other (Clarke & Garrett, 2004). Foreign accented speech is considered to an extreme form of accent variation due to the inclusion of unfamiliar phonetic pronunciations and non-native prosodic patterns of speech production that are caused by the non-native speaker mapping their L1 constraints onto their L2 production. In contrast, regional accent variation alters the production of speech in ways that should be more familiar to the observer (Floccia et al, 2006). Essentially, the production of various speech components may vary between regional accents, but each of those variations occurs within the constraints of the native language type (See Chapter 1, Section 1.7). The general assumption is that although both foreign and regional accent variation may influence speechreading performance, learning or adaptation to regional accent should be both easier and quicker.

Membership of a regional group is associated with the acquisition of certain speech traits (indexical properties) that are inherent to that region (Stuart-Smith, 1999). These traits influence the production of speech and are propagated through social imitation, producing a local population of people with an accent that is both distinctive and identifiable within a certain geographical area. The distinction between British regional accents varies along a continuum, with gradual changes occurring on a local geographic basis. Thus, two adjacent regions may be classified as having different accent types, yet still share many phonological similarities (Hughes, Trudgill & Watt, 2005). In terms of accent categorisation, this means that regions are associated with certain accent characteristics, but that these may be shared across regions or might vary within the region itself. For example within each region, variation associated with social class, together with individual variability relating to physiological factors, age and gender will have an impact on each individual's articulation of speech and thus the auditory and visual signals produced. Certainly previous studies in both auditory (Cox, Alexander & Gilmore, 1987) and visual (Kricos & Lesner, 1982) speech, have found that levels of intelligibility can vary between talkers despite no obvious accent differences (see Chapter 1, Section 1.5). These findings indicate that indexical factors other than accent can influence the production of speech. Thus, while an accent type can be said to contain various phonological factors, these will vary on a talker-by-talker basis.

There are five main accent regions within the UK and Ireland (see Chapter 1, Figure 1.8), the south of England, the north of England, Wales, the South of Ireland and Scotland and Northern Ireland (Hughes et al, 2005). Each of these regions is defined by a set of general pronunciation features that are shared within that region, but are distinguishable from the rest of the UK. For example, in Scotland and Northern Ireland the accent is rhotic, in that /r/ is pronounced at the end of a word and after a vowel. In comparison, in the majority of the South of England the accent is non-rhotic, so words such as 'bar' would be pronounced 'ba', with no emphasis on the /r/ at the end of the word (Hughes & Trudgill, 1987, see Chapter 1, Section 1.7 for full discussion).

Due to the regional accent variation mentioned above, individuals in the UK often come into contact with talkers whose accent differs from their 'home' accent. When this occurs they must attempt to deal with both phonological or phonetic variability and indexical variation, in order to comprehend the talker's message (Adank & McQueen, 2007). The majority of speech processing models include some facet of adaptation to variation in the speech signal, either as a process of active learning or a measure of habituation (Floccia et al, 2006; Norris McQueen & Cutler, 2003). Essentially, an unfamiliar accent interferes with speech comprehension due to its violation of stored expectations about prototypical speech production (Floccia et al, 2006). This disruption eventually alters the designation of recognised speech parameters in the speech processing system, incorporating the new information. Those newly adjusted boundaries are then used to facilitate comprehension of the auditory signal.

In terms of regional accent variation, a recent study (Evans & Iverson, 2003) found that listeners adjusted their categorisation of vowel sounds based on the accent of a carrier sentence. The authors suggested that the speech processing system formulates expectations about vowel production based on relatively short exposure to the accented speech. Furthermore, the stored exemplars representing speech sounds become weighted in favour of the accented speech, thus improving speech intelligibility. Adaptation can occur relatively quickly due to regional accent variation falling within the familiar parameters of native speech production. Essentially, listeners will be familiar with the phonetic productions used, albeit in a different context (Floccia et al, 2006). The suggestion therefore, is that the disrupting effect of an unfamiliar regional accent should be smaller than that of a foreign accent, since the foreign accent includes variations outside the expected parameters of the listener's stored lexical representations of speech.

Despite the above viewpoint, several auditory studies still observe a significant effect of regional accent upon auditory speech comprehension. For example, a recent experiment examined the effect of unfamiliar regional (Dutch) accent upon word processing speeds (Adank & McQueen, 2007). The study compared performance on a semantic categorisation (living / non-living) test and found that participants' reaction times were significantly increased when words were pronounced in an unfamiliar, as opposed to a familiar, accent. This result indicates that an unfamiliar regional accent can impair word recognition (Adank & McQueen, 2007). Further research has also indicated an adverse effect of regional accent upon lexical decision making (Floccia, Girard, Goslin & Konopczynski, 2006) and vowel identification (Labov & Ash, 1997) in spoken language tasks.

Unlike auditory speech perception there has been little comprehensive research to date on the effects of accent on visual speech perception. However, there are indications as to the potential effect of regional accent on speech recognition and understanding. A study by Ellis and colleagues (2001) noted that deaf participants from the North of England were significantly worse at matching pictures to visually presented spoken words than those from the South of England. The authors suggest that such poor performance by the northern English participants may have resulted from the group's lack of familiarity with the southern accent of the two talkers used in the study. However, two caveats apply. First, only the South of England accent was used as a stimulus and so no conclusions can therefore be drawn about the relative intelligibility of an accent from the North of England. Second, accent type was rather broadly defined and did not take into account sub-regional variations (see Chapter 1, Section 1.7).

To summarise, research indicates that a lack of familiarity with a regional accent has a detrimental effect on auditory speech perception, although the magnitude of the effect is likely to be smaller than that found for foreign-accented speech (Floccia et al, 2006). Visual research suggests a similar effect of accent (Ellis et al, 2001), though no firm conclusion can yet be drawn. Three experiments in Chapter 3 examine the effect of UK regional accent upon speechreading performance in greater detail to that shown previously. Experiment 3 was a questionnaire study, developed to identify difficult accent types, Experiments 4 and 5 investigated visual discrimination performance for different regional accents and measured the effect of regional accent upon speechreading performance.

3.1 Experiment 3: Questionnaire study on regional accent intelligibility and discrimination

The questionnaire study examined the potential effects of regional accent upon visual speech processing. By gathering the opinion of experienced speechreaders, the aim was to identify particularly difficult accent types that have a detrimental effect upon visual speech intelligibility. By examining the relationship between a participant's own regional location and the accents rated as difficult to understand, the results should allow us to determine whether familiarity with an accent type has any impact upon judgement of its intelligibility. Specifically, the experimental hypotheses were as follows:

- Familiarity with an accent type should influence judgements of accent intelligibility.
- A familiar accent type should be more visually discriminable than an unfamiliar accent type.
- There should be no clear relationship between speechreading ability and the influence of accent upon speech intelligibility.
- Visual speech intelligibility should be influenced by familiarity with an accent type as opposed to the actual characteristics of an accent type.

3.1.2 Method

Participants

172 hearing-impaired participants completed the questionnaire. Both male and female respondents were recruited, with an age range of 22 to 91 years (mean age 67). The participants varied in their hearing level; 16 were profoundly deaf, 65 were severely hearing impaired and 91 were moderately hearing impaired. All were recruited from English speechreading classes located in the East Midlands (Nottinghamshire, Leicestershire and Lincolnshire), the North of England (Lancashire) and the South of England (London, Norfolk, Dorset, Hampshire and Essex). The distribution of participants across those regions was as follows: Midlands = 19%, North of England = 9%, South of England = 72%. Recruitment of participants from sub-regions in Scotland was also attempted but was unsuccessful.

Materials

A short, 10-minute questionnaire was created utilising both closed and open questions (Appendix 3). Its brevity was designed to maximise return rate. Questions included:

- Demographic information (age, gender, self-reported hearing level, place of residence and place of birth)
- A measure of speechreading experience (number of years as a speechreader) and ability (self-reported)
- Reports on factors affecting speech intelligibility. These included regional accent type (pronunciation of speech), strength of accent (refers to the extent to which an individual speaker prescribes to the characteristics of an accent type – a strong accent would include all of the main facets of the accent type, a weak accent would include only some of the main accent factors), mumbling, speech rate, dialect (regional varieties of the native language) and audibility of speech production
- Reports on particular accent types which have a detrimental effect on speech intelligibility
- Reports on the accent types which are visually discriminable
- Reports on which aspects of certain accent types impair speech intelligibility

Procedure

230 questionnaires were distributed through contact with the English speechreading classes detailed above. The participants were mailed the questionnaires, or they were distributed within the speechreading class. Each respondent completed the questionnaire in their home environment. All

questionnaires were returned by post using a pre-paid envelope. 172 completed questionnaires were received in total, representing a 75% rate of return.

3.1.3 Results

Each participant was designated by his/her self-reported level of deafness (profoundly deaf, severely impaired and moderately impaired), his/her level of speechreading ability and his/her regional location (Midlands, South England and North England).

To determine whether accent had a negative effect upon speech comprehension, participants were asked to indicate their agreement with the statement "When conversing with others have you ever found that a person's accent can make it more difficult to understand what they are saying?" Responses were grouped initially by hearing level, with the majority of each group answering 'yes, sometimes' or 'yes, always' (88% of the profoundly deaf respondents, 92% of the severely impaired and 81% of the moderately impaired). This supports the claim that accent is an important factor which can disrupt visual speech intelligibility. When responses were re-grouped by reported speechreading ability, good, average and poor speechreaders responded equally to 'yes, sometimes' or 'yes, always' categories (65 – 71% of responses). Therefore, better speechreading ability does not appear to reduce the disruptive influence of accent.

Respondents were asked to rank accent type and strength together plus five other factors (fast rate of speech, slow rate of speech, mumbling, dialect and whispering) from 1 to 7 in terms of importance; 1 having the greatest negative impact upon speech comprehension and 7 having the least. Table 3.1 below summarises the responses for each of the different factors affecting speech intelligibility.

Factor / Rank	Accent Type	Accent Strength	Fast Rate of Speech	Slow Rate of Speech	Mumbling	Dialect	Audi-bility
1	7	13	25	6	39	9	29
2	10	11	23	4	27	14	20
3	15	14	28	7	11	14	12
4	12	29	4	11	5	15	12
5	20	16	5	12	2	22	8
6	26	10	7	10	7	19	9
7	11	8	8	51	9	8	10
Modal Res-ponse	**6**	**4**	**3**	**7**	**1**	**5**	**1**

Table 3.1: Percentage of participants giving each named factor a particular rank

Non-parametric testing using the Friedman rank test showed a significant effect of factor (x_2 (6) = 229.3; $p < 0.005$). From these responses it would appear that the majority of the respondents viewed mumbling (1) and audibility (1) as the two factors which had the most effect upon their understanding of speech, highlighting acoustic clarity of the speech signal as a major factor in intelligibility. Nevertheless, accent strength (4), dialect (5) and accent types (6) were all considered to exert some impact on speech intelligibility, albeit to a lesser degree.

In order to identify particularly unintelligible accent types, respondents were asked to give three examples of British accents they found difficult to speechread. The initial results are illustrated by Figure 3.1, which shows both regional accent types (colour coded) and the specific sub-regional accent types that were named by respondents in the questionnaire.

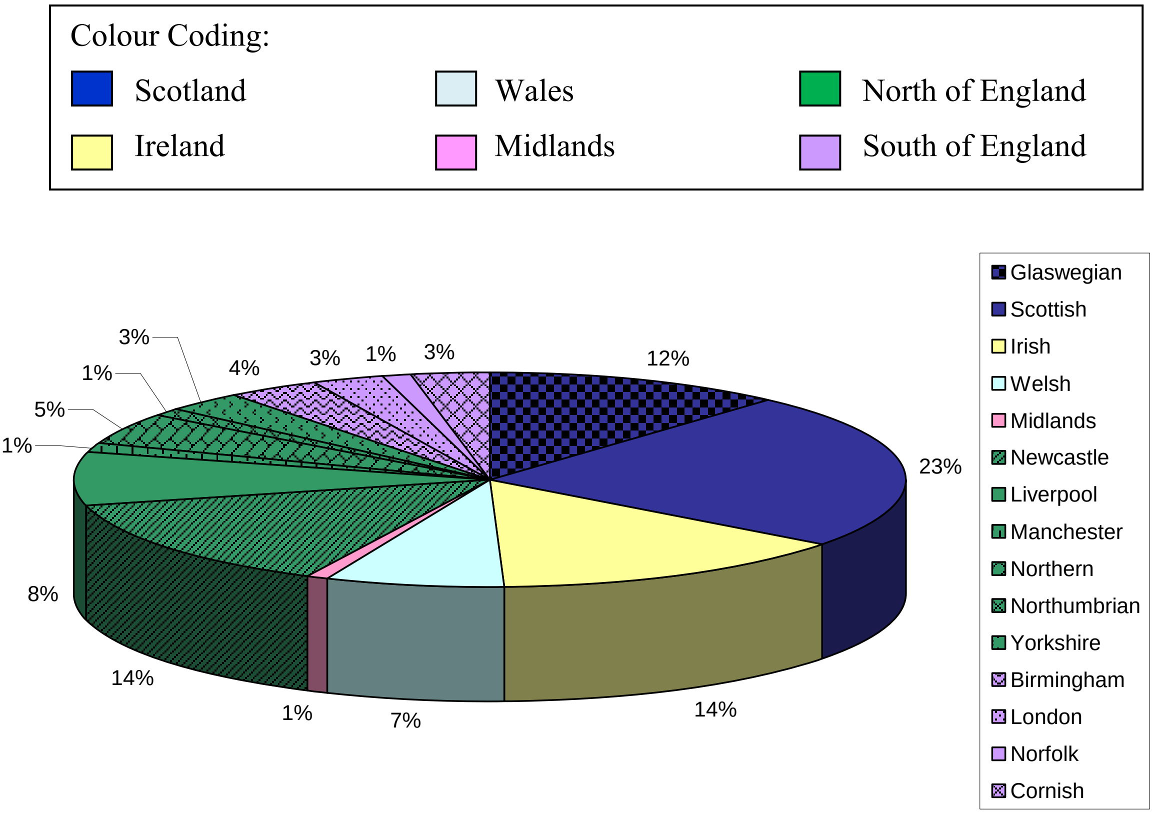
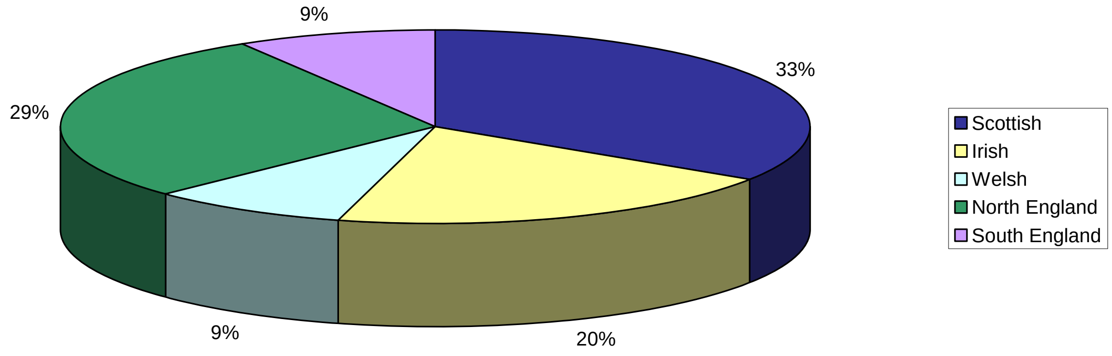


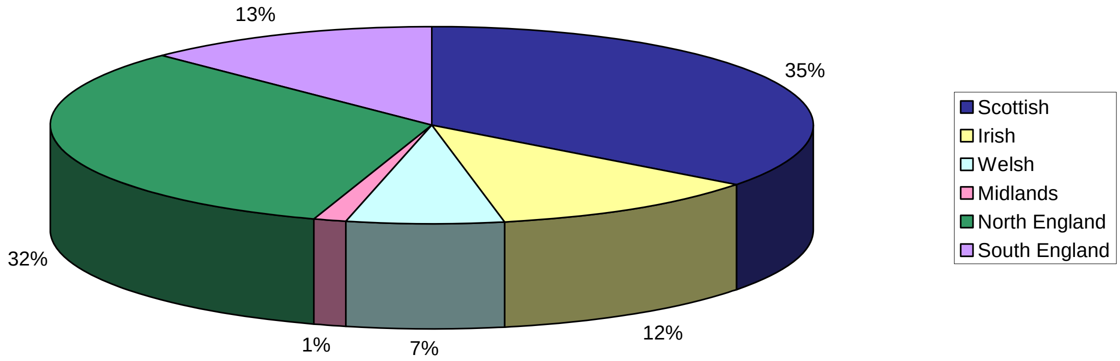
Figure 3.1: Percentage of regional accent types identified as difficult to understand

Figure 3.1 indicates that several accent types were considered unintelligible by speechreaders. The spread of results illustrates that the identification of difficult accent types was not restricted to one or two regions, since all five main accent regions were represented. However, considering that all of the respondents were English, there does appear to be an influence of familiarity upon accent intelligibility, because non-English accents (Scottish, Irish and Welsh) constituted the majority of the accents identified (56%).

Northern England Participants



Southern England Participants



Midlands Participants

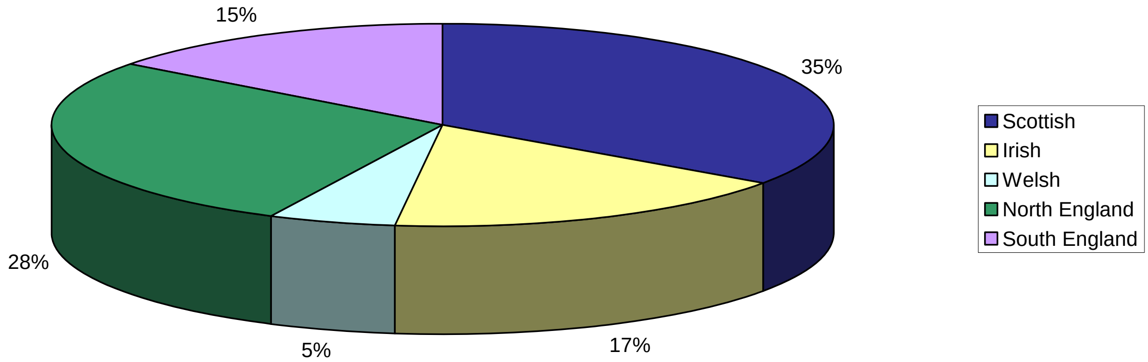


Figure 3.2: Each pie chart represents the recorded responses from participants born, and living in, Northern England, Southern England or the Midlands. The results are based on the percentage of respondents indicating that a particular regional accent type was unintelligible.

To further examine the relationship between speech intelligibility and accent familiarity, the results were re-plotted on a region-by-region basis, shown separately in Figure 3.2. There is little difference between the participant groups in Figure 3.2, indicating that certain accent types may be difficult to understand regardless of an observer's familiarity with them. For example, 28 - 32% of all three participant groups rated Northern English accents as unintelligible, including the Northern England participants themselves. If speech intelligibility was based on familiarity alone, we would have observed a different pattern of results for each region of participants, in which no group of participants would have rated their own accent as unintelligible. However, since Scottish, Irish and Welsh accents again make up the majority of accents identified (more than 50% in each case) it is possible that accent familiarity does have an influence, albeit on a different scale to that shown in auditory research. Essentially, sub-regional accent familiarity does not appear to influence the results, rather, a regional influence that is based on English versus non-English accents appears to be having an effect. This interpretation is not conclusive due to the lack of data from Scottish participants. Until the opposite pattern of results can be shown by non-English participants it is unsure whether these findings are due to familiarity or the characteristics of particular accent types. Finally, it was noted that very few participants identified the Midlands accent as difficult to speechread. This may be due to the Midlands accent having phonetic elements in common with both the North and South of England accent types, making the Midlands accent less distinctive and potentially less likely to cause intelligibility problems.

The visual discriminability of different accent types was examined by asking respondents to name three accents which they could distinguish using the visual modality alone. The results are illustrated by Figure 3.3.

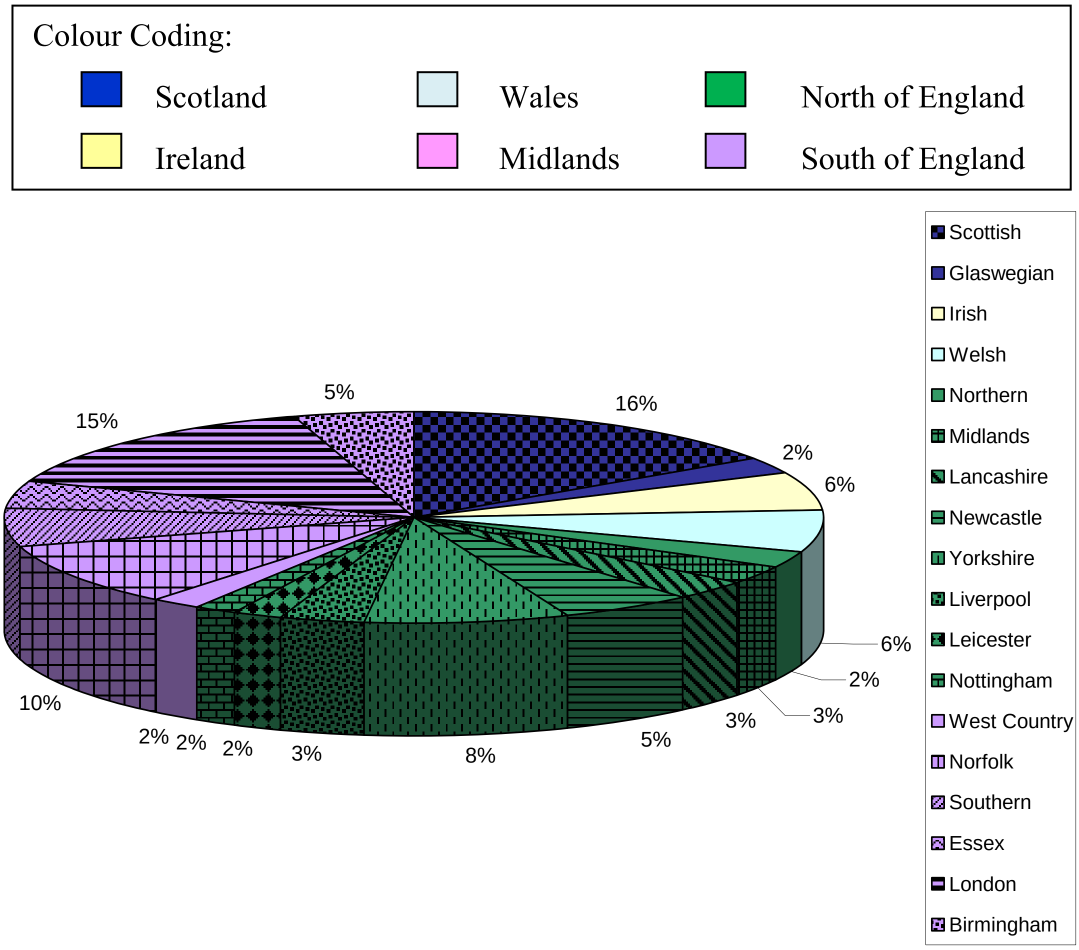


Figure 3.3: Percentage of participants stating accent type is discriminable through speechreading.

A comparison of Figures 3.1 and 3.3 highlighted a marked difference between visual distinctiveness and the impact of a regional accent upon speech intelligibility. Essentially, it is apparent that a greater proportion of English accents were reported as discriminable than were rated as unintelligible. This suggests that speechreaders might find it easier to discriminate familiar as opposed to unfamiliar accent types. Further analysis of the data contained in Figure 3.3 is reported in Table 3.2. The Table illustrates the percentage of participants from each regional location reporting a particular accent type as visually distinctive.

Accent Type / Place of Residence	Scotland	Ireland	Wales	Midlands	North of England	South of England
Midlands	13	8	8	19	21	31
South of England	16	4	5	3	20	68
North of England	19	19	0	13	13	38

Table 3.2: Percentage of participants naming regional accent type to be easily distinguishable, according to participant regional location

While the majority of participants found Scottish accents difficult to understand, those accents rated most visually distinctive were predominantly from the South of England. This suggests that the ability of a speechreader to discriminate an accent type may depend on their level of familiarity with it. First, when the data were more finely classified according to the sub-region (or even named cities) it was noted that local residents often named their own regional accent as most distinguishable. For example, it was noted that all of the respondents who acknowledged the Norfolk accent to be visually distinctive were located within the Norfolk region. In addition, 15% of participants from the South of England named the London accent as distinguishable, as opposed to only 6% of North of England participants and 11% of participants from the Midlands. Furthermore, 5% of Midland participants and 6% of North of England participants viewed the Nottingham accent as easy to discriminate as opposed to 1% of South of England participants. Thus, extensive experience of, and familiarity with, an accent type may aid in the discrimination of speech produced in that accent type. Conversely, a lack of familiarity with that accent may decrease the likelihood of successful accent discrimination.

Further analysis found a correlation between the self-reported ability to identify accent types through speechreading and the self-reported use of speechreading as a skill: $r(169) = 0.342$; $p < 0.01$, $r2 = 0.12$, 12% variance shared, and also with self-reported speechreading ability: $r(169) = 0.272$; $p. < 0.01$, $r2 = 0.07$, 7% variance shared. In other words, participants who reported a reliance on speechreading or good speechreading ability were also more likely to claim the ability to identify accent through visual speech. It is possible that both experience in the utilisation of the visual modality, and the ability to accurately speechread, are required before the more subtle differences in visible articulation become noticeable and then distinguishable as belonging to a particular accent set of articulations.

The remainder of the questionnaire was devoted to open-ended comments about aspects of accent variation that influence speech intelligibility. Comments focused around three central themes, articulation (68%) familiarity (20%), and expectation (12%). The majority of difficulties relating to articulation were ascribed to differences in vowel production (60% of articulation responses) e.g. “vowels have different shapes depending on accent” and “vowel lengths are changed". The remainder of responses falling under articulation centred upon specific accent characteristics, which again relate to the clarity of the visual signal. For example, "the internalisation of speech production; the Scots seem to swallow their speech and not move their lips much", and "Scottish speak quickly and use strange forms of words". In terms of familiarity, comments such as "Difficult to meet people outside of your area as you have not experienced their accent" and "Need to work out what accent is to get used to it – unfamiliar accents = different lip shapes", suggest that a lack of familiarity with an accent can reduce visual speech intelligibility. The remaining comments were grouped under expectations and relate to the problems associated with dealing with unexpected articulatory motion; "Difficulty in anticipating lip movement – the mouth movements are not as expected", and "Can't anticipate speed, content of conversation or how it's delivered". These comments indicate that not only are speechreaders aware of the difficulties caused by accent, they can also pinpoint which aspects of accent result in the greatest difficulty.

3.1.4 Discussion

Individual variation in the production of auditory speech has already been linked to intelligibility difficulties (Cox, Alexander & Gilmore, 1987). The aim of the present questionnaire was to examine the potential impact of regional accent upon visual speech perception, with a particular focus on intelligibility and discriminability. The results indicated that although accent was not considered the most significant cause of difficulties in visual speech intelligibility, it was considered to cause some difficulty regardless of hearing level or speechreading ability. Of particular interest was the response of the profoundly deaf participants; 88% considered accent as a source of speech intelligibility difficulties, highlighting accent as a factor that can influence visual, as opposed to auditory or audiovisual, speech intelligibility. This agrees with previous auditory research that has shown that an unfamiliar accent has a detrimental effect on auditory speech intelligibility (Floccia et al, 2006; Labov & Ash, 1997), indicating a potential similarity in the effect of talker-specific factors upon the perception of auditory and visual speech.

The next step in the analysis identified those accent types that were considered to have a detrimental effect on speech intelligibility. The majority of respondents considered Scottish and Northern English accents to be difficult to understand. These views were the same irrespective of a respondent's familiarity with an accent type, suggesting that the visual characteristics of an accent type have a greater influence upon intelligibility than familiarity with that accent type. This finding contradicts claims that auditory speech intelligibility is more related to familiarity with an accent type than to the characteristics of the accent itself (Floccia et al, 2006). This contradiction suggests that the perception of accented speech in the visual system may differ from the auditory system, with familiarity being of lesser importance in visual speech perception.

An alternative explanation is that some accent types may be more memorable or visually distinctive than others, leading respondents to note those accent types above other equally difficult but 'less memorable' accents. Moreover, familiarity may operate on a broader geographical scale than the sub-regional level proposed on the basis of auditory speech. Essentially, there may be a broad dichotomy between the Scottish, Welsh and Irish accents versus English accents. Accent familiarity may therefore still yield an effect on intelligibility of visual speech, albeit on a different scale of accent specification to that found using auditory speech.

In comparison to the results above, the majority of accents judged to be visually discriminable were the South of England accents. It would therefore appear that the accents considered visually distinctive are also those accents most associated with high intelligibility. Accent familiarity may also play a contributory role here, but the distinction between familiarity and intelligibility requires further systematic study. The pattern of respondent choices noted here suggests that experience with an accent type does determine those accents an observer is able to distinguish successfully, with high familiarity levels facilitating accent discrimination. Experiments 4A and B address the influence of familiarity on accent discrimination in a more systematic and laboratory-controlled manner.

Finally, comments on the causes of accent-related decrements in intelligibility highlight several areas of consideration – familiarity, articulation and expectation. Auditory literature highlights familiarity as the main cause of speech comprehension difficulties for accented speech (Floccia et al, 2006). The inclusion of familiarity as an influence on visual speech is indicative of some similarity between the auditory and visual speech processing systems. However, the majority of comments emphasised aspects of articulatory movement as opposed to familiarity, particularly vowel production. Much of the auditory literature that describes accent change also addresses the vowel length and pronunciation issues mentioned here (see Evans & Iverson, 2003). The present finding suggests that these accent-related alterations to articulation are visible to an observer and can affect their perception of the speech signal. Experiment 5 readdresses the issue of accent familiarity and visual speech intelligibility in a laboratory-controlled setting.

3.2 Experiment 4: A comparison of regional accent discrimination across the auditory and visual modalities

Accent provides more than mere variation in the production of a speech signal, it is also an indicator of birth-place, social position and is usually associated with certain stereotypes (Wells, 1982a). Accent forms a facet of speech categorisation, the usefulness of which depends on a listener's level of familiarity with, and depth of exposure to, various accent types (Wells, 1982a). There is substantial evidence that the more familiar a listener is with a particular accent the easier he/she finds it to discriminate, and to understand a talker speaking in that accent (Wells, 1982a).

Accent classification has been the focus of several studies within the auditory speech perception literature. One such study investigated the effect of a listener's native language type on regional accent discrimination (Ikeno & Hansen, 2006; Ikeno & Hansen, 2007). The study recruited three sets of participants; i) British, ii) American, iii) non-native (Chinese, Croatian, German and Japanese among others, all of whom were non-native English speakers). Participants were asked to discriminate between three UK accents produced by talkers from; i) Cambridge, ii) Cardiff and iii) Belfast, using auditory words, phrases and sentences. British participants were significantly more accurate (90%) at discriminating the UK accents than the non-native (55%) or American (72%) participants. Thus, the authors conclude that in order for accurate regional accent discrimination to occur, the listeners must be native speakers of the language with a contingent knowledge of, and familiarity with, the potential regional accent variation. The authors further suggest that the acoustic cues that denote a particular regional accent may be too subtle for the inexperienced or untrained listener. A similar suggestion has been put forward in the visual research domain, where Soto-Faraco and colleagues (2007) reported that the ability to discriminate language type from the visual signal was constrained by linguistic experience. Essentially, only those observers who were a native speaker of one of the two languages used (Spanish or Catalan) were successful at discriminating between them. Non-native speakers performed at chance (see Chapter 1, Section 1.6 for full description). This provides the basis for Experiment 4B where the ability of non-native speakers of English to distinguish between regional accents will be tested.

Experiment 4A investigates the auditory and visual discrimination of regional accent, together with the influence of regional accent upon visual and auditory speech intelligibility. Experiment 2 has already demonstrated that foreign accent can be discriminated using the visual modality. However, the cues for discriminating regional accent may be less visually distinctive than those denoting a foreign accent (Floccia et al., 2006). Cues that may used to identify an auditory accent include patterns of pronunciation, phonemic distribution and rhythmical characteristics (Wells, 1982a). Rhythm falls within the category of prosodic features, which also includes factors such as stress patterns, intonation and voicing (Wells, 1982a). However only pronunciation, stress patterns and rhythm are likely to be conveyed by the visual signal, since the other acoustic characteristics, such as voicing, are more difficult to discern visually. Thus, observers are likely to use speech rate (Berger, 1972), viseme production (Kricos & Lesner, 1985), rhythm (Munhall et al., 2003) and extent of mouth opening and closing (Tye-Murray &

Folkins, 1990) to distinguish visually between accent types. By comparing participants' performance on auditory and visual accent discrimination, we can test the hypothesis that the visual cues, which are likely to be less numerous and distinctive than those in the auditory modality are not as effective at signalling accent.

The second general aim of Experiment 4A was to provide a preliminary examination of the effects of regional accent upon visual and auditory speech intelligibility. It was expected that both modalities would be affected by accent type, with both the characteristics of the accent type and an individual's familiarity with that accent influencing their ability to understand a talker's speech.

Experiment 4B examines the effect of native language type upon regional accent discrimination. Essentially, Ikeno and Hansen (2006) posited that the ability to discriminate regional accents depends on the listener's native language type and their contingent experience of the accents which they are asked to discriminate. A listener who speaks English as his/her L2, or who lives in America with little experience with British accents, will find it more difficult to discriminate between British regional accents than a listener who is native to the UK. By recruiting non-native speakers of English for Experiment 4B, we could determine if a similar effect exists in visual accent discrimination. The experimental hypotheses were as follows:

- Experiment 4A: regional accent discrimination should be possible based on visible articulations alone, though accent discrimination utilising the auditory signal is expected to be more successful.
- Experiment 4A: there should be a positive correlation between speechreading accuracy and performance on the visual accent

discrimination task. Essentially, proficiency in comprehending the visual signal should give observers an advantage in extracting accent cues.

- Experiment 4A: based on the questionnaire results, the Glaswegian accent should have a detrimental effect on both visual and auditory speech intelligibility.
- Experiment 4B: non-native participants who have little knowledge of British regional accents should be significantly poorer at both auditory and visual accent discrimination of regional accents than native English speakers from Britain.

3.2.1 Experiment 4A Method

Participants

Ten participants, both male and female, were recruited for the study with an age range of 18 to 48 years old (mean age: 31). Participants were native English speakers who had been born in England and had lived in the East Midlands area for a minimum of 5 years. Each one reported normal vision and hearing.

Stimuli

A total of 168 BKB (Bench, Kowal & Bamford, 1979) sentences were recorded in 12 sets of 14 sentences. Sentences were chosen as the stimulus type in order to ensure that observers were exposed to the full set of linguistic cues (phonetic and prosodic) for accent (see Floccia et al., 2006 for similar methodology). This represents a more ecologically valid measure of accent effects than the utilisation of speech segments (such as words or syllables). Consequently every experiment in this thesis uses sentence length materials. Each set of sentences in the present experiment was produced by two talkers; one with a Glaswegian accent, one with a Nottingham accent, for counterbalancing purposes. The talkers used for the recordings were grouped as follows: six talkers (three male, three female) with Midlands (Nottingham) accents, six (four male and two females) with Scottish (Glaswegian) accents. In order to ensure that each accent type was fairly represented by the six talkers from each group, each talker had to meet certain criteria. The Glaswegian accent was rhotic and exhibited vowel shortening, as detailed by Aitkens Law (Wells, 1982b). In other words, a vowel was phonetically short unless it was followed by a voiced fricative or /r/. The Glaswegian talkers also omitted the phoneme /ʊ/, leading to the homophones 'pool' – 'pull', 'full' – 'fool' etc. Each talker showed a tendency to use a glottal 't' or /ʔ/, altering words such as 'butter' to 'buʔʔer'. Finally, the Glaswegian accent was characterised by having a single phoneme /ɔ/ common to 'lot' and 'thought' and producing homophones such as 'cot' – 'caught' (Wells, 1982b). The Nottingham accent was non-rhotic and included the phoneme /ʊ/, though not the phoneme /ʌ / leading to the homophones 'put' – 'putt'. The Nottingham talkers also exhibited /ɒ/, differentiating between the vowel sounds of 'lot' /ɒ/ and 'thought' /ɔ/. Furthermore, the Nottingham talkers were unlikely to use a glottal 't' and exhibited phonemic vowel lengthening. These two regional accents were chosen because they differed both visually and auditorially. There was also a local availability of participants who were familiar with the Nottingham accent. IHR links to the regional section in Glasgow also provided access to participants familiar with the Glaswegian accent, a requirement for Experiment 5.

Each recording featured a talker's face, fully illuminated using three high power lamps placed at right angles to reduce shadowing. The recordings were made against a neutral white background with only the face and neck of the talker visible. In every recording, the talker’s full face was filmed from a camera (Sony Digital Camcorder, DSR-200AP) placed 1.5m directly in front. Each recorded sentence was preceded by 1s of the talker’s static face in a closed mouth position, and followed by a further 1s of static footage again in the closed mouth position. Each sentence was recorded using a microphone, with the auditory signal later separated from the visual signal to allow both unimodal and bimodal stimulus presentations.

Procedure

Each testing session contained 168 sentences; 84 visual and 84 auditory. The experiment was divided into two phases; 1) visual and 2) auditory.

Phase 1 (visual modality): The initial phase of the experiment utilised 84 of the recorded visual sentences in order to produce a measure of i) word identification and ii) accent discrimination for that modality alone.

Task i): Each participant was seated at a table directly in front of the computer screen and instructed to watch each video clip carefully. They were presented with 60 visual sentences, comprised of 12 sets of 5 sentences, each set spoken by a different talker and all presented in random order. The sentence materials were counterbalanced across participants by making two recordings of each set – one produced by a Nottingham talker, the other by a Glaswegian talker. The talker viewed by each participant for that set was then alternated so that half would view the set produced by a Nottingham talker, the remainder viewing a

Glaswegian talker. This type of counterbalancing was repeated across all of the stimuli used in this experiment. They were instructed that the talker would utter one sentence per video clip. Their task was to identify the spoken words in a sentence and type their response on the computer keyboard.

Task ii): Participants were then asked to attempt to discriminate each talker's accent type based on a further 24 visual sentences, two sentences per talker. They were asked to view each video clip and then indicate whether they thought the talker had a Glaswegian or Nottingham accent by typing in '1' for Glaswegian and '2' for Nottingham. The sentence materials were counterbalanced as before.

The two tasks of speech intelligibility and accent discrimination were blocked to remove the possibility of a reduction in performance due to cognitive set switching (Rogers & Monsell, 1995).

Phase 2 (auditory modality): The same two tasks were presented in the auditory modality (sentence identification and accent discrimination), with the same procedure used except that the participants listened as opposed to viewed the stimuli.

3.2.2 Experiment 4A Results

For the word identification task, the results represent the number of keywords correctly identified within a sentence set. Every sentence had three keywords, the total potential score for each complete set was 180 keywords, 90 keywords for each accent type. For the accent discrimination task, the results represent the percentage of correct responses per accent type. There were 12 sentences per accent type and so 50% correct represents six correct responses out of 12.

Regional accent discrimination was greater than chance in both the visual: $t(9) = 2.57$, $p < 0.05$ and auditory: $t(9) = 15.57$, $p < 0.05$ modalities, as illustrated by Figure 3.4. As expected, performance for the auditory task (87% correct) was more accurate than for the visual task (58% correct) and this difference was significant; $t(9) = 6.34$, $p < 0.05$. Thus, regional accent was easier to discriminate when speech was presented as an auditory, rather than a visual, signal. Finally, there was no correlation in the participants' ability to classify visual and auditory accent types: $r(10) = 0.109$, $p > 0.05$, indicating that the two discrimination tasks were not linked in terms of performance.

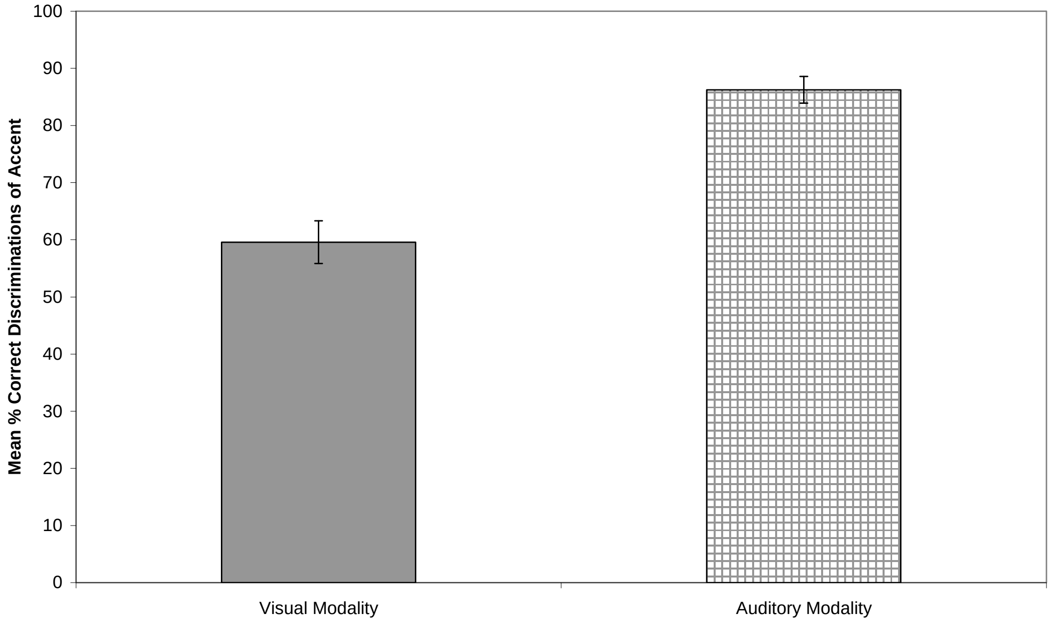


Figure 3.4: Mean percent correct discrimination of accent based on speech presented in the visual or auditory modality, shown here with standard error bars

Performance on the word identification task was significantly affected by regional accent type, as illustrated by Figure 3.5 below. The adverse effect of the

Glaswegian accent upon performance was significant in both modalities; auditory: $t(9) = 6.83$, $p < 0.05$, visual: $t(9) = 2.43$, $p < 0.05$. Thus, the Glaswegian accent is distinct from the Nottingham accent in both sound and appearance and is also significantly less intelligible.

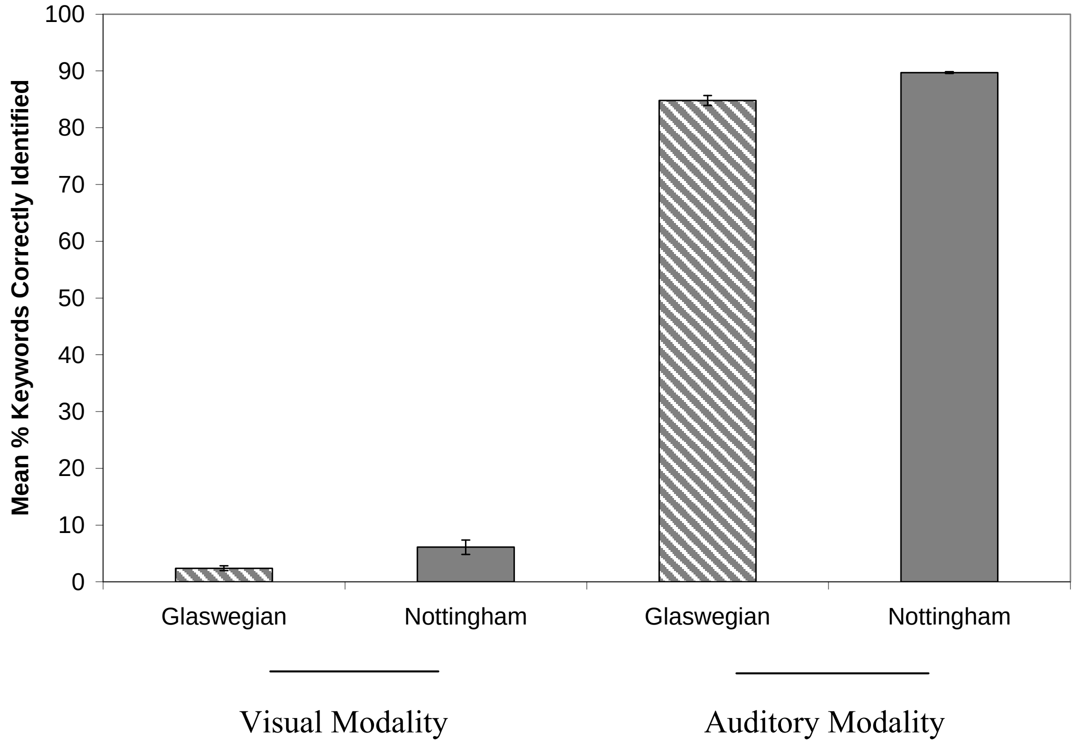


Figure 3.5: Mean % keywords correct for visual and auditory sentences spoken with either a Glaswegian or Nottingham accent, shown with standard error bars.

There was no correlation in performance between the visual and auditory speech perception scores for either accent: Glaswegian: $r(10) = .265$, $p > 0.05$, Nottingham: $r(10) = .40$, $p > 0.05$. This indicates that the ability to identify words in the visual or auditory modality is not necessarily linked. However, this result should be viewed with caution due to possible floor effects resulting from the low speechreading scores illustrated by Figure 3.5. The usual spread of speechreading scores using the BKB sentences is between 10 and 50% correct (MacLeod & Summerfield, 1987). The present spread of scores is between 2 and 10% correct, which is substantially lower. There are two potential explanations for the low performance rates. First, the talkers were chosen on the basis of their acoustic accent characteristics as opposed to their relative speechreadability. For this reason, various idiosyncratic talker factors, such as lip shape and speech rate could be influencing intelligibility. Second, previous research (Yakel et al, 2000) indicates that using multiple talkers, as we have done here, reduces speechreading performance. Thus the number of talkers used in the experiment may have had an adverse effect on the participants' ability to speechread.

Finally, to determine whether there was a link between visual accent discrimination and speech intelligibility, the visual accent discrimination and speechreading scores were correlated separately for each accent type: Glaswegian accent: $r(10) = 0.290$, $p > 0.05$, Nottingham accent: $r(10) = 0.355$, $p > 0.05$. However, there was no significant relationship.

Experiment 4, part B

Previous auditory research on regional accent discrimination suggests that to discriminate British regional accent types accurately, the listener must be a native speaker of English (Ikeno & Hansen, 2006). The authors further suggest that performance also depends on relevant language experience, with a native English speaker from Britain more likely to successfully discriminate British accent types than a native English speaker from America. This indicates that without implicit knowledge of regional accent characteristics, a listener will be unable to accurately discriminate between accent types at the same level as a native speaker (Ikeno & Hansen, 2006). Experiment 4A showed that native English speaking participants with a high level of familiarity with British regional accents were able to discriminate between a Nottingham and Glaswegian accent presented in both the visual and auditory modalities. In order to determine whether these results were contingent upon experience with the English language and British accent types, the accent discrimination task was repeated with 10 non-native speakers of English to test the predictions made by Ikeno and Hansen (2006). The hypothesis was that participants who were non-native speakers of English would be significantly less accurate at British regional accent discrimination than native English speakers.

3.2.3 Experiment 4B Method

Participants

Ten participants, both male and female, were recruited with an age range of 21 to 40 years (mean age: 29.2). All were fluent English speakers who had been born outside the UK (Poland, Germany, Greece and France) and had been living in the East Midlands for less than 2 years.

Stimuli

The lists of 24 BKB (Bench, Kowal & Bamford, 1979) sentences used in the accent discrimination task of Experiment 4A were again used here.

Procedure

The procedure was the same as that described in Experiment 4A for the accent discrimination task, presented in both auditory and visual modalities.

3.2.4 Experiment 4B Results

The results are presented here in direct comparison with Experiment 4A in order to allow us to compare native versus non-native participant performance. Analysis was conducted using a two-factor (2 modality x 2 language experience) mixed factorial ANOVA. A significant effect of modality was found: $F(1, 18) = 66.288$; $p < 0.05$. Thus participants were significantly better at distinguishing accent type when the stimuli were auditory (mean: 79% correct) as opposed to visual (mean: 58% correct). There was also a significant effect of language experience: $F(1, 18) = 6.747$, $p < 0.05$. Thus, the native English speakers were better at discriminating accent (mean: 73% correct) than the non-native speakers (mean: 65% correct). There was a significant interaction between modality and language experience: $F(1, 18) = 4.737$, $p < 0.05$. This interaction indicates that the effect of language experience differs significantly across the two modalities used in the task, as illustrated by Figure 3.6. Post-hoc analysis using parameter estimates found a significant difference between the native and non-native participants for auditory ($t = 3.729$, p. < 0.01), but not visual ($t = .569$, p. > 0.01), accent discrimination. Thus, the previous auditory accent effects (Ikeno & Hansen, 2006) are reliable and survive replication. In contrast, comparable visual accent effects were not found.

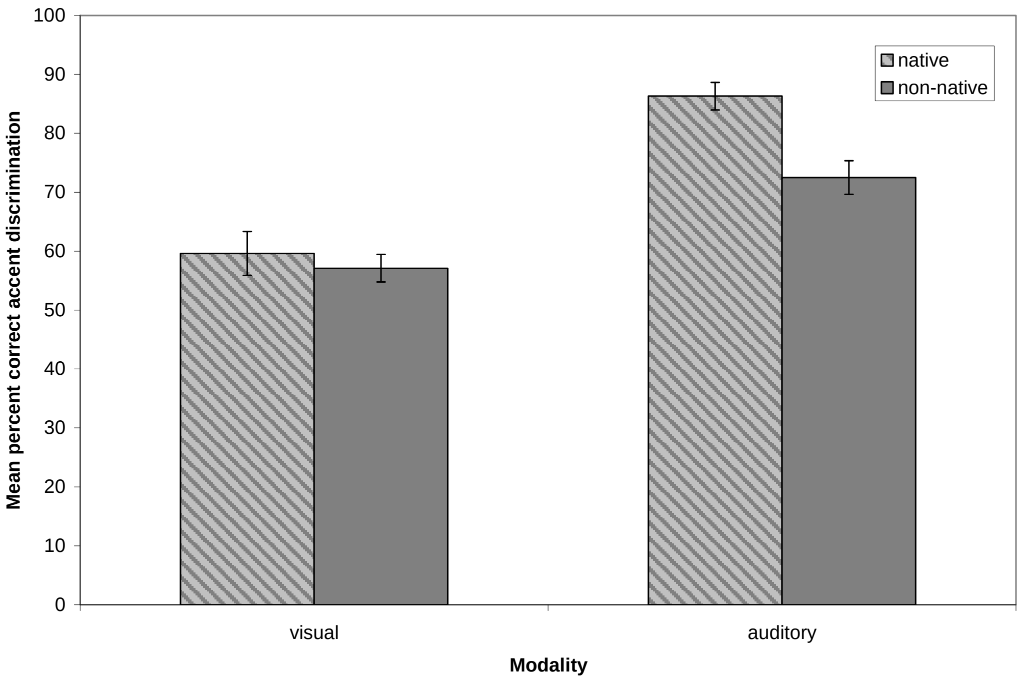
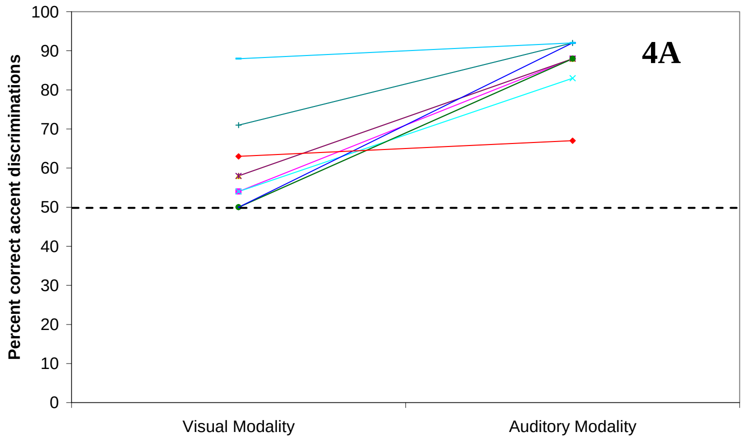


Figure 3.6: Mean percent correct discrimination of accent based on visual or auditory sentences by native and non-native English speaking participants

The overall impression from Experiments 4A and 4B was that while linguistic experience appeared to be a constraining factor in auditory accent discrimination, it had less of an impact on accent discrimination in the visual modality. Interestingly, when the data were plotted across the individual participants, it became apparent that although all listeners were capable of discriminating accent from the auditory stimulus at a level above chance, not all observers were able to successfully discriminate accent type from the visual signal (see Figure 3.7). The majority of participants from Experiment 4A exhibit a clear advantage for auditory as compared to visual accent discrimination. In comparison, there is a larger spread of data within the non-native participants discrimination scores (4B) for both modalities. There are several potential explanations for the data shown in Figure 3.7, first, all participants may be more successful at decoding auditory rather than visual cues as to accent type. Second, the non-native participants may have adopted different strategies in order to compensate for distinguishing accents from their L2, producing the wider spread of data. However, although these strategies may have led to high levels of individual variability within Experiment 4B, overall non-native participants were successful at the task with performance for both the auditory ($t(9) = 7.867$, $p < 0.05$) and the visual modality ($t(9) = 3.042$, $p < 0.05$) being significantly above chance (50%).



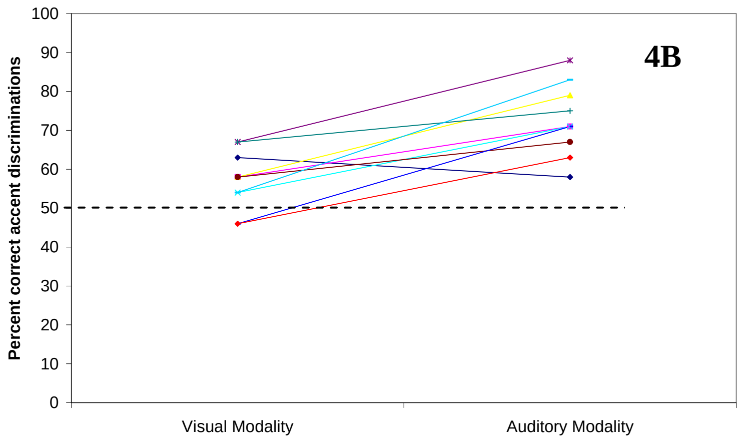


Figure 3.7: Discrimination performance by individual participants. Colours indicate individual participants, each experiment used different participants and so the colours used are not consistent across 4A and 4B.

3.2.3 Discussion

The results from Experiments 4A and 4B support the claim that regional accents are discriminable on the basis of visual information alone. This overall pattern is the same as that seen for foreign accent in Experiment 1. Thus, whilst auditory research indicates that non-native accents represent an extreme form of speech variation (Clarke & Garrett, 2004), it would appear that the visual differences between regional accents are still distinctive enough to allow reasonably accurate discrimination.

The regional accent types used here were chosen for the number of distinct phonological and prosodic differences between them, differences which have been shown previously to impact auditory speech intelligibility (Nathan, Wells & Donlan, 1998; Arnold & Hill, 2001). The assumption was that the accents, which constitute variation between two main regional accent types, should be easier to discriminate than accents with fewer differences between them, such as those contained within an accent sub-region (see Chapter 1, Section 1.7). It is entirely possible that two accents more alike in structure and realisation, such as the Nottingham and Sheffield accents (both from the Midlands region), would be more difficult to discriminate visually. Further research would be necessary to investigate such a possibility.

Large differences between the visual and auditory modalities were observed with accent discrimination being significantly better for auditory speech. This was perhaps unsurprising due to the different nature of auditory and visual accent cues. Auditory cues to regional accent include nasality, voicing, articulation, rhythm, stress, phonetic and phonemic differences (Wells, 1982a). The cues relating to accent in terms of the visual signal are fewer in number, and include visible articulation, rhythm, stress and viseme differences. This informational asymmetry may explain the recorded performance differences.

Auditory research suggests that native speakers of a language have more of the knowledge necessary for accurate regional accent discrimination within that language than non-native speakers (Ikeno & Hansen, 2006). Native language speakers should therefore be better at discriminating regional accents than non-native speakers. This suggestion is consolidated by visual research which indicates that linguistic experience is a constraining factor in language discrimination using visual speech (Soto-Faraco et al., 2007). The results reported in Experiment 4B indicate that non-native English speakers were significantly less accurate at discriminating British regional accents presented in the auditory modality than native English speakers. Interestingly, this effect was not replicated for visual accent discrimination ($p > 0.05$). This null result could indicate that both native and non-native participants were utilising similar processing strategies to decode visual accent cues. Alternatively, the lack of an effect could be due to the potential differences in processing strategies between the visual and auditory systems, or it may simply be due to floor effects. The results from the present experiment are insufficient to distinguish between these possibilities.

Experiment 4A enabled a direct test of the prediction, made on the basis of the questionnaire results reported in Experiment 3, that there would be a relationship between speechreading performance and visual accent discrimination. The present results do not support that prediction since no significant correlation was found, though it should be noted that this null result may be due to floor effects. However, supporting evidence for the suggestion that accent discrimination does not require speechreading proficiency is present in studies of language discrimination in pre-linguistic infants (Weikum, Vouloumanos, Navarra, Soto-Faraco, Sebastian-Galles & Werker, 2007). The authors found that infants as young as 4 months old were able to visually discriminate between the languages of French and English. Although the task differs from that examined here, the results are indicative of the infants' sensitivity to language and foreign accent variations in the visual signal, a sensitivity that does not arise from speech comprehension due to their lack of linguistic ability. The conclusion therefore is that the visual signal carries sufficient dynamic articulatory information to indicate accent type, without the need for speechreading proficiency to provide awareness of potential linguistic variation.

Interestingly, although speechreading accuracy is not necessary for accent discrimination, it appears that the visual cues for accent type do influence speechreading performance. This is indicated by the detrimental effect of the Glaswegian accent upon both auditory and visual speech intelligibility, as observed in Experiment 4A. This objective measure provides support for the self-reported comments in Experiment 3 and is the first demonstration of specific regional accent effects upon visual speech intelligibility. Experiment 5 was developed to examine this effect in further detail by attempting to repeat the detrimental effect found here with a larger pool of participants recruited from several regional locations.

3.3: Experiment 5: Regional accent familiarity and its effect on speechreading performance

Experiment 5 further examined the effects of regional accent on speechreading ability. The main aim was to separate the effects of regional accent type from the effects of accent familiarity on speechreading performance. The effect of accent familiarity upon speech intelligibility was reviewed in Chapter 1, Section 1.7, where a lack of familiarity with an accent was shown to have a negative effect on auditory speech intelligibility in children (Nathan, Wells, & Donlan, 1998; Nathan & Wells, 2001), adults (Labov, 1989; Labov & Ash, 1997; Munro & Derwing, 1995) and on speech-in-noise intelligibility (Clopper & Bradlow, 2006). In a series of experiments, Floccia and colleagues (2006) investigated the effect of French regional accents on the processing of (French) auditory speech using a lexical decision-making task. The recorded outcome indicated that speech produced in an accent that was unfamiliar to the participant resulted in slower reaction times, suggesting that speech processing is more effortful when an unfamiliar accent is encountered. The pattern of deficits was found to depend on the geographical region from which the subject came, linked to geographical location and associated exposure to accents. The authors suggested that it was the level of exposure to an accent that influenced speech processing, rather than any specific acoustic characteristic of the accent itself. Familiarity with an accent type should therefore be a determining factor in the intelligibility of accented speech.

These accent effects may be explained through either the abstractionist or the encoding theory of speech perception (Nygaard & Pisoni, 1998, for a full description see Chapter 1, Section 1.3). Abstractionist, or normalisation, theory states that to understand the message spoken by different speakers the observer must account for and remove the variation between their own speech and that of the talker. This is a cognitively demanding process possibly resulting in phoneme mapping errors and reduced comprehension. Alternatively, the encoding approach suggests that the specific characteristics of a talker's speech pattern are encoded in long-term memory, producing an associated short drop in performance whilst cognitive resources are used (Pisoni, 1997). However, once an observer is familiar with a talker's articulatory characteristics, perception of that individual's speech should improve until understanding has risen to previous levels (Pisoni, 1997). Thus, familiarity with, and consequential adaptation to, an accent type, should improve speech intelligibility. This process of adaptation can occur through active learning or the screening of variation through normalisation.

Familiarity with a talker has been shown previously to positively influence the intelligibility of visually presented words (Lander & Davies, 2008). The study consisted of five sections, these were split into three test sessions containing the word lists, and two 'training' sessions where an audiovisual presentation of a two minute story was produced by a single talker. There were three conditions, i) familiarised same, where participants viewed the same talker for all five sections (test and training), ii) familiarised different, where the talker used for the three test sessions was different from the talker producing the two training sessions and iii) control, where participants completed a puzzle between each test session. The results showed that speechreading accuracy improved to a greater extent in the familiarised same condition (mean improvement = 18%) than in the familiarised different (improvement = 6%) or the control condition (5%). The authors suggest that increased exposure to a talker allows observers to adapt to the speaking style of that individual, hence improving their ability to speechread that talker (Lander & Davies, 2008). This theory of talker familiarity could be extended to encompass regional accent familiarity for stimuli presented in the visual modality. The main hypothesis is that familiarity with the visual correlates of a regional accent type should produce an advantage when attempting to speechread talkers with that accent. Conversely, lack of familiarity with an accent should have a detrimental effect.

Experiment 5 addresses the effect of regional accent and accent familiarity on visual speech comprehension by comparing speechreading performance across two distinct accent types (Glaswegian and Nottingham) and three sets of participants (from Nottingham, Glasgow and Southampton). These accent types were chosen because of their distinct visual differences, as shown in Experiments 4A and 4B. Experiment 4A demonstrated an impairment in speechreading performance for sentences presented visually in the Glaswegian accent compared to the Nottingham accent. However, in Experiment 4A participants were all resident in the Nottingham region and so accent familiarity might equally explain the results observed. In Experiment 5, participants were recruited from three locations in order to more directly investigate the issue of accent familiarity. Two groups were highly familiar with one of the accents used. Nottingham participants were familiar with the Nottingham accent and Glaswegian participants were familiar with Glaswegian accent. It should be noted that it is difficult to precisely specify levels of familiarity. Most people will come into contact with a variety of accents both through social contact and through other mediums, TV for example. However, it was assumed that their familiarity with their own 'home' accent would be greater than that with any other accent type. The Southampton group provided a control measure. While they had no specific familiarity with either regional accent they could be expected to have knowledge of certain aspects of the Nottingham accent, since there are some shared characteristics between South of England and Midlands accents. For example, both accents are non-rhotic and exhibit phonemic vowel lengthening (Wells, 1982b)

Two further considerations for Experiment 5 were speechreading proficiency and talker familiarisation. First, it is possible that better speechreaders are more able to cope with regional accent variation and with unfamiliar accents than poor speechreaders, due to their proficiency in extracting speech information from the visual signal. Second, both auditory and visual research has indicated that prolonged or repeated exposure to a talker improves speech intelligibility for that talker, potentially compensating, to a certain extent, for any regional accent effects on performance (Floccia et al, 2006; Lander & Davies, 2008). On the basis of this second consideration an element of continued exposure was included in Experiment 5 through the utilisation of block-by-block variation in talker and accent to determine if a short period of exposure would be sufficient to improve performance levels. The experimental hypotheses were thus:

- The Nottingham and Glaswegian participant groups should show a specific advantage for visual speech produced in the 'home' accent, most familiar to them.
- The Southampton participant group should show some advantage for the Nottingham accent, though this may be smaller than that exhibited by the Nottingham participant group.
- Better speechreaders should be more able to compensate for accent variation than poor speechreaders.
- Continued exposure to a particular talker and accent type should improve speechreading performance over time.
- Some level of variation in speechreadabilty should be found between talkers within the two accent groups, but this effect on performance should not be greater than the observed accent effects.

3.3.1 Method

Participants

Fifty-eight participants were recruited for the study, all were native English speakers and reported good hearing and normal (or corrected to normal) vision. The participants were allocated to three groups dependent on regional location. In total, 24 participants were recruited from Nottingham, 17 from Glasgow and a further 17 from Southampton. All participants had lived in the region from which they were recruited for a minimum of four years. Furthermore, each participant had been born within that region, thus Nottingham participants were born in the East Midlands, Glaswegian participants were born in Central Scotland and Southampton participants were born in the South of England. This criterion ensured their familiarity with the accent of that region (their 'home' accent).

Stimuli

The 12 talkers used for the recordings were those described in Experiment 4A, with six Glaswegian speakers and six Nottingham speakers. Forty sentences were recorded for each talker. Each recording featured a talker's face, fully illuminated using three high power lamps placed at right angles to reduce shadowing. The recordings were made against a neutral white background with only the face and neck of the talker visible. In every recording, the talker's full face was filmed from a camera (Sony Digital Camcorder, DSR-200AP) placed 1.5m directly in front. Each recorded sentence was preceded by 1s of the talker's static face in a closed mouth position, and followed by a further 1s of static footage again in the closed mouth position. Each clip contained only visual information.

A total of 260 sentences from the BKB set (Bench, Kowal & Bamford, 1979) were recorded. These formed three lists. List 1 was the basic speechreading measure comprised of 20 sentences produced by a talker with a Southern English accent. This test was always carried out first and featured the same set of sentences, presented in random order, for every participant. The further 240 sentences were split into 12 sets of 20 sentences. Every set was spoken by a different talker. In order to counterbalance the 240 sentences across participants, the sentences were split into list 2 and list 3 (120 sentences each). Half of the participants viewed list 2 spoken by Glaswegian talkers and list 3 spoken by Nottingham talkers, the remaining half of the participants viewed the opposite. This removed the possibility of sentence effects upon the results.

Procedure

Each participant was seated at a table directly in front of the computer screen. They were instructed that a talker would utter one sentence per video clip, which they were asked to watch carefully. Their task was to identify the spoken words and type their response on the computer keyboard. They were not required to understand the entire sentence and any word that was typed in was scored.

The experiment consisted of two phases. In Phase 1, each participant completed list 1, the basic speechreading measure. In Phase 2, each participant completed the speechreading test comprised of lists 2 and 3, presented in a particular order that was predefined according to one of two group memberships. These two groups addressed the issue of short-term adaptation via repeated exposure to a particular talker and accent type. Group 1 viewed the sentences in a fully random order with the talker changing sentence by sentence. Group 2 viewed the sentences in 12 blocks, with a change of talker after every 20 sentences. The order of the talkers for Group 2 was randomised within accent type. So the participant would view six talkers from one accent group and then six talkers from the second accent group. This was to determine if a continued exposure to an accent type would improve performance for that accent. Group 1 comprised of 12 Nottingham, 9 Glaswegian and 9 Southampton participants, Group 2 comprised of 12 Nottingham, 8 Glaswegian and 8 Southampton participants.

3.3.2 Results

Each participant's score was generated using a loose keyword scoring system with errors in morphology ignored. Every sentence contained three keywords with a point awarded for each correctly identified keyword. In this case, the potential total of keywords for the basic speechreading measure was 60. The two lists of sentences in the main body of the experiment contained 360 keywords (i.e. 3 x 120).

A similar level of performance was recorded for all three participant groups on the speechreading measure (Nottingham participants: 13% keywords correct, Glaswegian participants: 11% keywords correct, Southampton participants: 14% correct), indicating a similar baseline of speechreading ability within each group.

The results were analysed using a two-factor (2 accent type x 3 participant location) mixed factorial ANOVA. The mean scores for the two observed accent types were found to differ significantly: $F (1, 55) = 75.618$, $p < 0.01$. Thus the intelligibility of the Glaswegian talkers (mean: 3% keywords correct) was significantly worse than the Nottingham talkers (mean: 7% keywords correct). There was no significant effect of participant location: $F (1, 55) = .132$, $p > 0.05$, indicating no overall difference in speechreading performance across the three participant locations; Glasgow (mean: 5% correct), Southampton (mean: 6%

correct) and Nottingham (5% correct). There was a significant interaction between accent and participant location: $F(1, 55) = 3.276$, $p < 0.05$. This interaction is illustrated in Figure 3.8 and indicates that the effect of accent type differed significantly across participant location.

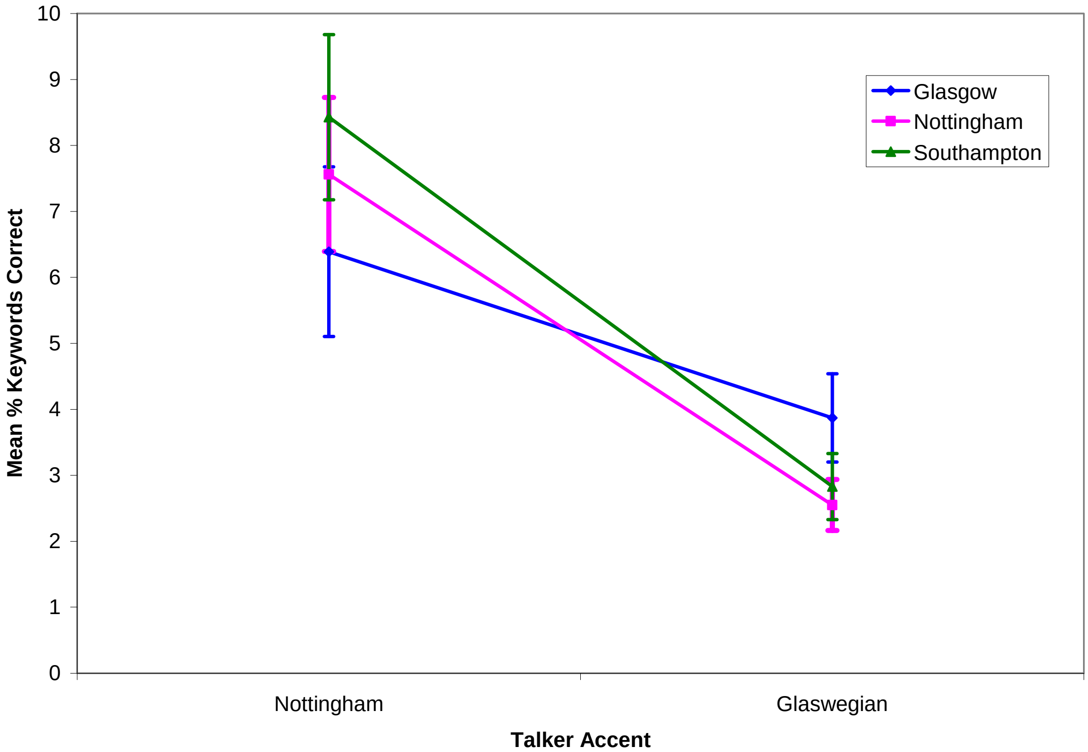


Figure 3.8: Mean % of keywords correctly identified from sentences spoken with either a Nottingham or Glaswegian accent by participants from Glasgow, Nottingham and Southampton, shown with standard error bars

Further analysis of the interaction used estimated marginal means and 95% confidence intervals. The results indicated that although the three participant groups exhibited similar performance for the Nottingham talkers, it was the performance scores for the Glaswegian talkers that showed a significant difference across participant location (see Table 3.3). Essentially, it appears that the Glaswegian participants were better at speechreading the Glaswegian talkers than the Nottingham participants, as exhibited by the lack of crossover between the scores of those two groups. This finding suggested a possible small effect of accent familiarity upon performance, supported by a post-hoc t-test that showed the difference in performance between the Nottingham and Glaswegian participants for the Glaswegian talkers to be significant ($t(39) = 1.829$, $p < 0.05$, one-tailed).

Participant Location	Talker Accent Type	Mean % Correct	95% confidence lower bound	95% confidence upper bound
Glasgow	Nottingham Glaswegian	6.4 3.9	3.7 2.8	9 5
Nottingham	Nottingham Glaswegian	7.6 2.5	5.3 1.6	9.8 3.5
Southampton	Nottingham Glaswegian	8.4 2.8	5.8 1.7	11.1 3.9

Table 3.3: 95% confidence intervals for Glaswegian, Nottingham and Southampton participants for talkers with a Glaswegian or Nottingham accent

The next issue to consider was the effect of continued exposure to a particular talker and accent type. The analysis compared the performance of participants who viewed the talkers varied on a trial-by-trial basis with those who viewed the talkers on a block-by-block basis with contingent familiarisation effects. Table 3.4 illustrates the initial results.

Talker Accent:	Nottingham		Glaswegian	
Talker Presentation Style / Location	Trial-by-trial	Block-by-Block	Trial-by-trial	Block-by-block
Nottingham	25	29	9	10
Glasgow	25	19	13	15
Southampton	32	28	12	8

Table 3.4: Mean speechreading performance across participant groups for trial-by-trial and block-by-block variation.

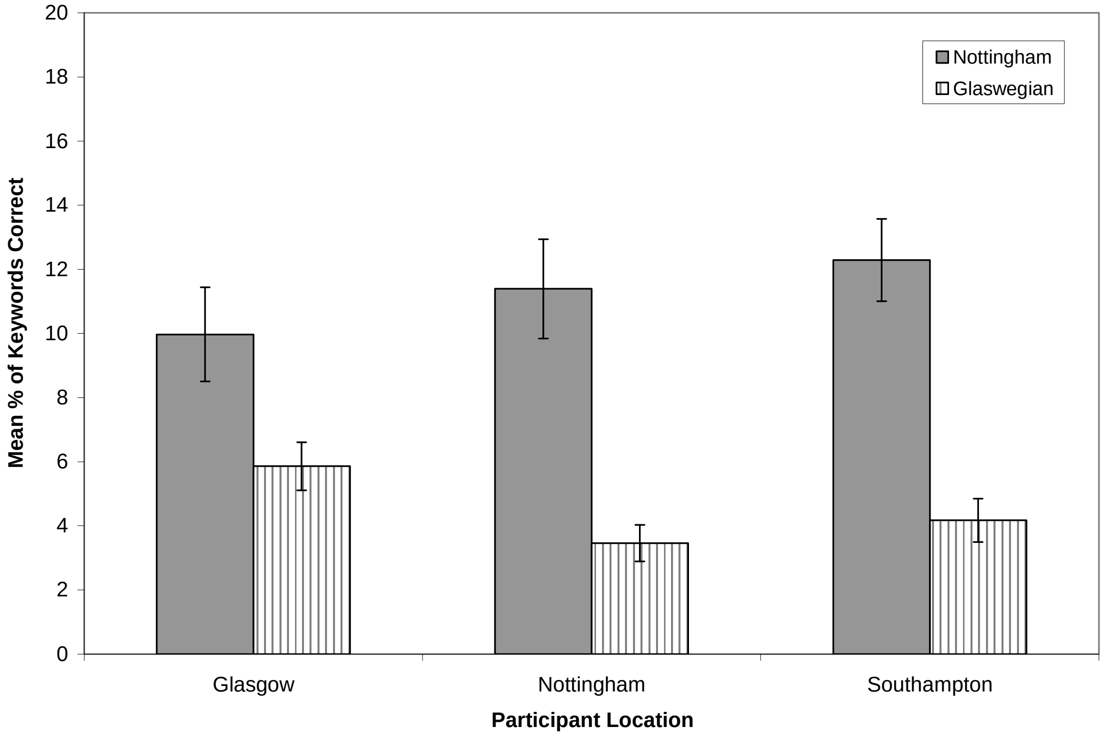
A two-factor (2 accent x 2 presentation type) mixed factorial ANOVA was used. No significant effect of presentation type was found: $F (1, 56) = 114.030$, $p > 0.05$ and there was no significant interaction: $F (1, 56) = .006$, $p > 0.05$, between the factors. The single significant result was that of accent type ($p < 0.05$). Therefore, relatively prolonged exposure (the blocks were 20 sentences in length) to a single talker did not rapidly improve performance. It is possible that an even longer block would perhaps produce an advantage, or that larger subject groups would have increased the significance of the results, but, based on the present results, it would appear that varying talkers on a block-by-block basis does little to improve overall speechreading accuracy or reduce the effect of accent variation on speechreading performance.

To address the issue of speechreading proficiency, the participants were split into groups of ‘good’ and ‘poor’ speechreaders by way of a mean split on the basis of their performance on the basic speechreading measure (list 1). Mean performance was 10% and 28 participants had scores at or above 10% and so were designated ‘good’ speechreaders, the remaining 30 participants (< 10%) were designated ‘poor’ speechreaders. The mean percentage of keywords correctly identified by each group is illustrated by Figure 3.9.

The ‘good’ and ‘poor’ speechreaders were adversely affected by the Glaswegian accent across all three subject groups. Analysis of both groups performance using a two-factor (2 speechreading proficiency x 2 accent type) ANOVA found that the performance values for the good and poor speechreaders differed significantly; $F(1, 56) = 50.039$, $p. < 0.05$. There was also a significant interaction between accent type and speechreading proficiency; $F(1, 56) = 27.028$, $p. < 0.05$, indicating that proportionally there was a greater effect of accent type for the good rather than the poor speechreaders, i.e. the good speechreaders were more impaired by the Glaswegian accent.

Further analysis using one-way ANOVA showed that in the case of the ‘good’ speechreaders, the Glaswegian participants were significantly more accurate when speechreading the Glaswegian talkers than the two English participant groups; $F(2, 25) = 3.45$, $p. < 0.05$. There was no significant effect for the Nottingham participants’ speechreading the Nottingham-accented talkers. Neither was there a significant difference in performance across the two accent types in the poor speechreaders group.

‘Good’ Speechreaders



‘Poor’ Speechreaders

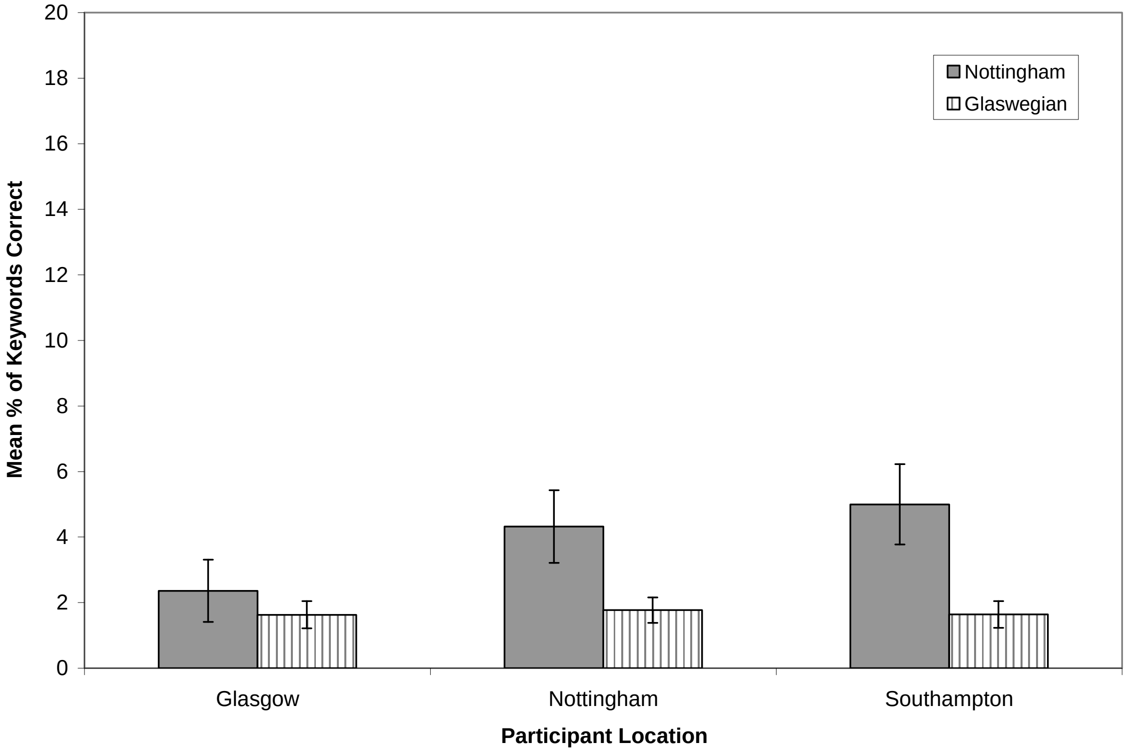


Figure 3.9: Mean percentage of keywords correctly identified by ‘good and ‘poor’ speechreaders across accent type, separated by participant location and shown with standard error bars

The final question examined was the effect of talker variability upon the general accent effects. Figure 3.10 illustrates the speechreading performance of the three subject groups, plotted on a talker-by-talker basis.

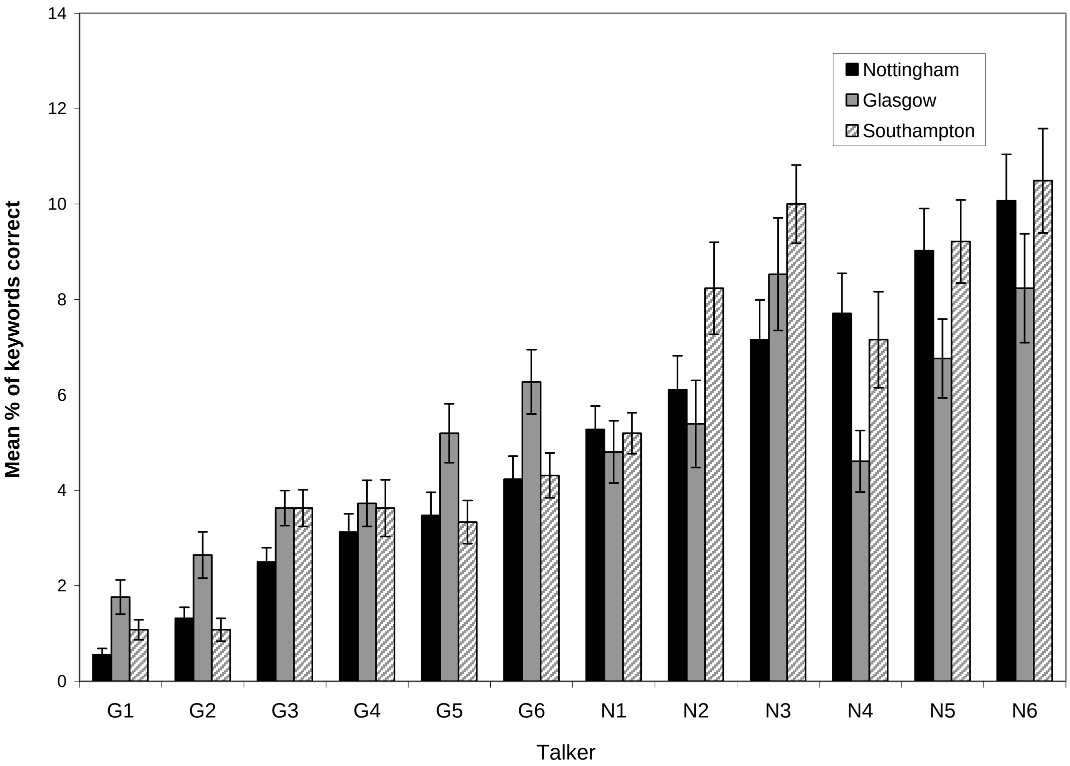


Figure 3.10: Mean percent keywords correct for each talker (Glaswegian accent = 'G', Nottingham accent = 'N'), across participant location, shown with standard error bars.

Figure 3.10 illustrates that there was some variation both between the talkers and across the subject groups, but in general all three sets of participants found the Nottingham talkers easier to speechread than the Glaswegian talkers. The pattern of performance suggested that there may be some significant variation between the talkers from both accent groups. Using repeated measures ANOVA (12 levels) the results were analysed for each group of participants. The analysis showed that the mean scores for each of the twelve talkers differed significantly for each participant group (Nottingham: $F(4, 84) = 16.379$, $p < 0.05$, Glasgow: $F(5, 78) = 6.461$, $p < 0.05$, Southampton: $F(4, 60) = 16.243$, $p < 0.05$). Post-hoc multiple pairwise comparisons using the Bonferroni threshold correction method found that these effects were distributed differently in the three groups (for full table of effects for the Nottingham and Glaswegian participants see Appendix 4). Both the Southampton and Nottingham groups were similar in that the majority of significant differences between talkers came from between the two accent groups. The only exception was talker G1 who was less visually intelligible than the other talkers in the Glaswegian group. In comparison, the Glaswegian participants showed a pattern of variation both within the Glaswegian group of talkers *and* between the two groups of talkers. Essentially, Glaswegian talkers G5 and G6 were significantly more intelligible than both the other talkers within the Glaswegian group and Nottingham talkers N1, 2 and 4 for the Glaswegian participants.

3.3.3 Discussion

The results reported here for visual speech differ from those reported for auditory speech on a number of points. Within auditory research, lack of familiarity with an accent type was found to yield the greatest detrimental effect on auditory speech intelligibility, over and above any effect associated with the acoustic properties of the accent (Floccia et al, 2006). Comparatively, the results reported for the present study indicate that the opposite is true within the visual modality for the Nottingham and Glaswegian accents. Thus, although familiarity does have an impact, the main effect of accent relates to the visual characteristics of an accent type. Essentially, the characteristics of the Glaswegian accent were such that even observers who were extremely familiar with the accent type

(Glaswegian participants) exhibited poor performance for the Glaswegian talkers. These results, combined with the small, yet significant, advantage shown by the Glaswegian participants for their own accent when compared to the performance of the Nottingham participants, suggest an inter-relationship between accent familiarity, the visual characteristics of an accent type and speechreading performance.

Auditory research on accent has previously indicated that an observer must compensate for a talker's idiolect (Hughes, Trudgill & Watt, 2005), or speech characteristics (indexical information) in order to perceive speech. This normalisation, or encoding of talker indexical information, (Nygaard & Pisoni, 1998), occurs over a period of time whilst the unfamiliar speech sounds are registered and understood, resulting in a decrease in intelligibility for speech produced by an unfamiliar talker or in an unfamiliar accent (Floccia et al, 2006). Our results partially support the suggestion that a similar effect may occur in visual speech processing. Certainly the majority of the participants' initial speechreading performance was hampered by the Glaswegian accent, indicating that the acoustic aspects of accented speech also influence the visual signal. However, as detailed above, it seems unlikely that this effect was based purely upon accent familiarity, but rather on a relationship between familiarity and the characteristics of an accent type. Whether this effect can be mitigated through further repeated exposure, providing observers with the opportunity to learn the characteristics of the Glaswegian accent, is a question that will be investigated in Chapter 5.

The pattern of results presented in Experiment 5 suggests a complex relationship between accent and intelligibility. First, the Glaswegian participants were adversely affected by the Glaswegian accent, which suggests that there are certain elements of that particular accent type which negatively alter the visual signal, making it more difficult to discern. The specifics of those elements are outside the scope of this study, but possibilities include the shortened vowel structure prescribed by Aitkins Law (Wells, 1982a) which may make segmentation of visual speech more difficult through an associated faster tempo of speech and consequential increased co-articulation. There is also the lack of the phoneme /ʊ/, which makes homophones of words such as pull and pool, full and fool and so on. This increases the likelihood of confusions between words potentially leading to a reduction in speechreading accuracy. These aspects of the Glaswegian accent may make it more difficult to interpret the speaker's message, even when familiar with the accent type. Second, although familiarity with the Glaswegian accent did not produce performance equivalent to the Nottingham accent, it did appear to improve accuracy levels to some degree in comparison to the English participants, particularly among the better speechreaders. The implication of this is that some of the difficulties associated with the Glaswegian accent can be compensated for through familiarity, whilst others appear resistant.

A surprising result was that a higher level of speechreading ability, as shown by those observers labelled 'good' speechreaders, did not improve speech perception of the Glaswegian talkers to a level on a par with the Nottingham talkers. This indicates that improved comprehension of the visual signal does not wholly compensate for those alterations caused by an unfamiliar or visually difficult accent type. This result can be related to the research conducted by Labov and Ash (1997), who found that, although increasing levels of context (phrase, then sentence) improved the comprehension of accented speech, context alone was not always enough to compensate fully for the detrimental effects of an unfamiliar accent. Indeed, once a particular vowel sound had been incorrectly identified by a subject unfamiliar with the accent, they seemed averse to altering their perception of it, even when the preceding phrase made their assessment of the sound meaningless. This indicates that although context may improve performance it may not fully compensate for accent-related misperceptions. It follows that although good speechreaders in the present experiment were able to decipher more of the accented visual cues, this did not help them to 'fill in the gaps' in the remainder of the sentence due to confusions arising from alterations to the visual signal caused by the accent. That is, their expectations of how speech should appear may have led them to interpret the accented visual signal incorrectly, producing the observed poor performance for the Glaswegian talkers. This suggests that even practiced speechreaders may find an unfamiliar accent difficult to process when they initially encounter it. However, it should be noted that the overall speechreading performance levels recorded for Experiment 5 were poor when compared to previous studies using the BKB sentences (MacLeod & Summerfield, 1987). As such the results may be subject to floor effects potentially due, as previously mentioned, to the utilisation of multiple talkers or to idiosyncratic talker features rendering the talkers difficult to speechread in general. Further research using a different set of talkers, or utilising a set of highly proficient speechreaders, would be necessary to determine the cause of the poor performance levels.

Finally, there was no difference between the trial-by-trial and the block-by-block variation of the stimuli, indicating that the potential familiarisation to both a particular talker and a particular talker's accent type in the blocked condition did not offer any significant advantage to the observer compared to the trial-by-trial variation condition. This is in direct disagreement with auditory research, which has found a significant improvement in the processing of an unfamiliar accent when the test items are presented in a blocked order (Floccia et al, 2006). The authors postulated that this improvement came about through adaptation to the unfamiliar signal. Thus, as the participant's exposure to the unfamiliar accent increased, a shift in their stored representations of speech should occur to incorporate the phonetic content of the new accent, facilitating speech perception (Floccia et al, 2006). The disparity between the present results and those reported by Floccia et al (2006) are indicative of a potential difference in the abilities of the visual and auditory speech processing systems to adapt to different accent types. This will be further examined in Chapter 5. The results of the present study simply suggest that the switch from talker to talker is as cognitively demanding in the block-by-block condition as it is in the trial-by-trial condition. However, it should be noted that each block was only 20 sentences in length, whereas the blocks used by Floccia and colleagues (2006) were 32 sentences in length. It is possible, therefore, that a longer block may have had more of an impact upon the results.

Finally, the results indicate that talker variability was present within each accent group. This was unsurprising in view of previous research which has shown the relative comprehensibility of talkers with similar accents to be vastly different (Kricos & Lesner, 1982; Lesner & Kricos, 1981). Indeed, the research by Floccia et al (2006) also found differences in the relative comprehensibility of the two talkers they used to represent each accent category, indicating that talker variability occurs for both auditory and visual speech. It can only be assumed that, similar to Kricos and Lesner's (1982) research, the production of visemes and thus each talker's intelligibility, varied on an individual basis within each of the accent groups due to an amalgamation of many factors, the specifics of which are outside the scope of this thesis. However, it is worth mentioning that the Glaswegian participants exhibited greater variability in their performance when speechreading the Glaswegian talkers than did the English participants. This could be indicative of their familiarity with the accent type interacting with aspects of talker variability, improving the intelligibility of some talkers above others. The English participants exhibited no such effect, suggesting that the accent type of the talkers was the most important aspect for observers unfamiliar with the accent type. The inter-relationship between talker and accent variation is considered in more detail in Chapter 4.

3.4 General Summary

To conclude, the results shown in Chapter 3 represent a first step towards quantifying the effect of regional accent upon visual speech processing. Experiment 3 showed that deaf speechreaders consider regional accent an important aspect of talker intelligibility. Experiment 4 illustrated that regional accents are discriminable using both the auditory and visual modalities. Experiment 5 indicated that there exists a complex relationship between familiarity with a regional accent and the actual visual articulatory gestures of the accent type itself. Familiarity with a regional accent type appears to give observers an advantage but does not fully compensate for any derived paucity in the visual signal, indicating the importance of accent as a factor in the speechreadability of a talker.

Chapter 4. Accent as a facet of talker variability

'Variety is the very spice of life, that gives it all its flavour'

William Cowper

The perception of speech requires a listener to extract meaningful linguistic information from a signal that is subject to a high level of talker-related variability (Goldinger, Pisoni & Logan, 1991). This variability is caused by an amalgamation of many different factors, including speech rate (Berger, 1972), accent type (Lesner, 1988) and voicing differences (Bradlow, Nygaard & Pisoni, 1999). A talker's 'idiolect' (Wells, 1982a) is their unique speaking style formed through their choice of dialect and grammar and shaped by each of the factors listed above. In Chapters 2 and 3, I showed that a talker's accent type is an important aspect of their idiolect, influencing the intelligibility of visual speech. However, other talker-specific influences were also apparent, highlighted by the talker variability evident within each regional accent group (Chapter 3, Figure 3.9). The purpose of Chapter 4 is to compare the effects of accent and talker variability upon visual speech with the aim of determining whether variation in the signal caused by a change in accent between talkers will exacerbate the effects of talker variation upon visual speech intelligibility.

The general view in auditory speech processing is that talker-specific information is encoded alongside the linguistic aspect of a message. These 'indexical' (Nygaard & Pisoni, 1998) properties of a talker's speech carry information about the talker's identity, their background and region of origin together with cues about their emotional state. The indexical properties of speech production also influence the talker's articulation of speech sounds which in turn affects the listener's ability to interpret linguistic information. Speech perception therefore involves the parallel processing of linked linguistic and indexical information. The consequence of this is that when multiple sets of talker information form the input to the speech processing system, the required utilisation of cognitive resources to encode the constantly changing indexical information can result in reduced comprehension of the linguistic aspects of speech (Nygaard & Pisoni, 1998).

The shift from talker to talker has been shown to have a detrimental influence upon vowel recognition (Assman, Nearey & Hogan, 1982), same-different judgements of acoustic letters (Cole, Coltheart & Allard, 1974) and the identification of words in noise (Mullenix, Pisoni & Martin, 1989, see also Sommers, Nygaard & Pisoni, 1994). These findings indicate that the negative influence of talker variability on the auditory modality reliably affects many aspects of speech perception. To highlight the nature of that effect we shall examine one study in more detail. Mullenix and colleagues (1989) required participants to identify a word presented in background noise, at a level of 70 dB. Half of the subjects listened to words produced by a single talker, the other half to a list produced by 15 different talkers, all of who had a Midwestern American dialect and accent type. Performance for the single-talker list (40.6% correct) was significantly more accurate than for the multiple-talker list (33.9% correct). The authors suggested that this detrimental effect was due to the sensitivity of the auditory processing system to acoustic variability at the initial stages of low-level auditory processing. This sensitivity could result in a perceptual deficit when the parameters of the auditory signal are consistently altered by talker factors. Alternatively, talker variability may result in an interaction between talker characteristics and linguistic information (Mullenix et al, 1989). Essentially, the encoded properties of a talker's mode of speech production may interfere with subsequent lexical analysis of a different talker's speech within the multiple-talker list. However, the specific nature of this interference was not determined by their results.

Despite this evidence of a detrimental effect of talker variability, there is also a potential advantage of such detailed processing of the speech signal and associated indexical information. Goldinger et al (1991) found that recall of word lists was more accurate for the initial items in a multiple-talker list presented at a slow rate (one word every 4000 ms) than a single-talker list. This advantage was reversed when the presentation rate was increased (one word every 250 ms) and also dissipates for words presented later in the list, as illustrated by Figure 4.1. The effects, shown by the impact of a variable upon items presented early within a list, demonstrate the influence of talker variation upon rehearsal effects. Essentially, the recall of words from the beginning of a list represents the efficiency of rehearsal, or encoding, of those items – a 'primacy effect' (Martin, Mullenix, Pisoni & Summers, 1989). The influence of a variable upon items later in a list, in comparison, is considered to affect short-term memory only (Martin et al, 1989). Therefore in this case the authors suggest that talker variation in the acoustic speech signal affects not only initial processing, as indicated by the detrimental effect of talker variability upon speech intelligibility, but also talker variation remains an integral component of speech representations in long-term memory (Goldinger, Pisoni & Logan, 1991). This additional talker-specific information appears to be detrimental to recall when a task is demanding (fast presentation rate) but acts as a positive influence when detailed processing can occur (slow presentation rate, primacy effects). These findings, together with those of Mullenix et al (1989) confirm the suggestion that the effects of talker variability upon auditory speech perception and recall are a result of the processing time and cognitive resources required to encode talker-specific information.

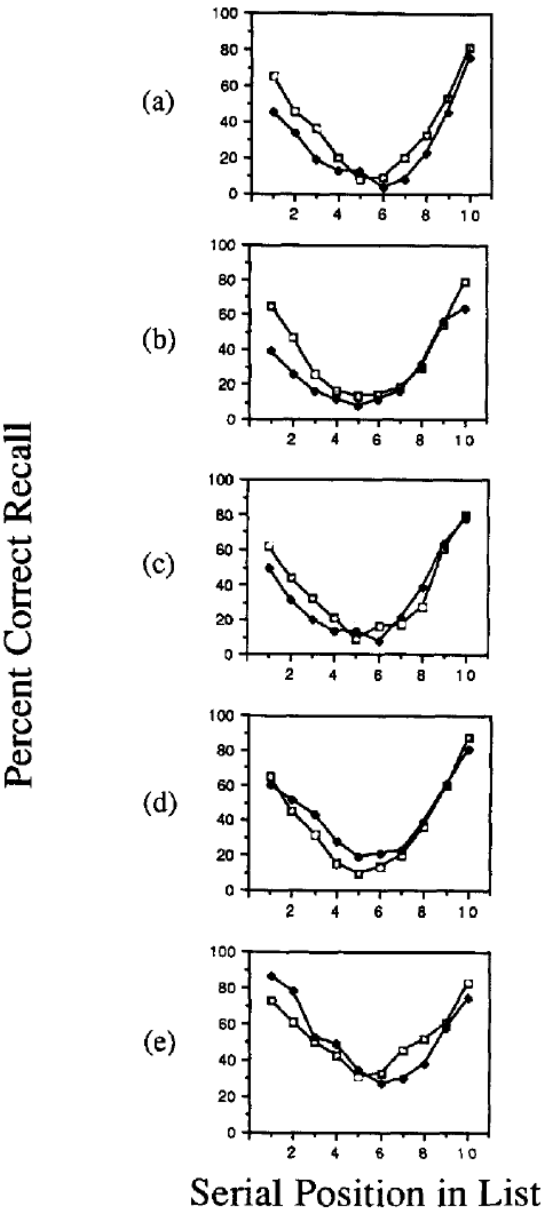


Figure 4.1: Mean percentages of correctly recalled words for both the single- and multiple-talker lists as a function of serial position and presentation rate, collapsed across word confusability. The five panels display the results at each rate of presentation, one word every (a) 250 ms, (b) 500 ms, (c) 1,000 ms, (d) 2,000 ms, and (e) 4,000 ms. Open squares represent single-talker lists; filled squares represent multi-talker lists (Goldinger et al, 1991).

Similar effects have also been recorded within the visual speech domain, the most influential study being an examination of the effects of talker variability upon speechreading performance (Yakel, Rosenblum & Fortier, 2000). Speechreading performance was compared using single- and multiple-talker productions of BKB sentences (Bench, Kowal & Bamford, 1979), visually presented. The results showed that performance was significantly more accurate for single-talker lists (55.8% keywords correctly identified) than multiple-talker lists (47.9% keywords correct). Interestingly, when the experiment was repeated using colour as a superficial difference, with performance compared across a single talker with one colour tint applied to the videos and a single talker with multiple colour tints varying on a trial-by-trial basis, performance was not affected. This indicates that variability which does not influence the actual visible articulation of a talker does not have an impact upon speech intelligibility. Thus, talker variation effects in visual speech should only be produced by factors that directly alter the informative cues in the visual signal, such as accent type and speech rate.

Several studies have demonstrated that talkers vary in their production of visual speech (Kricos & Lesner, 1982; 1985). This variation has a direct influence upon the intelligibility of talkers, with viseme categories and error rates changing on a talker-by-talker basis (Demorest & Bernstein, 1992). This variation relates to both physiological and linguistic differences between talkers, but also includes variation caused by the lack of defined articulatory movements to produce certain speech sounds (Lesner, 1988). For example, stop consonants such as the alveolar stop /t/ and the velar stop /g/ involve the complete constriction of air flow through the mouth (Carr, 1999). However, the exact point of closure can vary on an individual basis (Lesner, 1988). Similarly, the production of a bilabial stop such as /p/ can be produced with either pursed or flat lips (Lesner, 1988). The result of this is that the articulatory movements used by each individual talker can vary along parameters other than accent and speech rate, defined only by the individual's chosen manner of speech production.

Research has sought to define visual intelligibility by measuring the influence of various talker characteristics. For example, an examination of lip shape and the degree of lip-rounding on the visual intelligibility of vowels (Montgomery & Jackson, 1983), found that the general size of lip opening could be used to predict (i.e. explains over 40% of the variance in vowel confusions) talker intelligibility. However, the degree of success in lip opening at predicting speechreadability was found to be talker dependent in that, of the four talkers used, only three were successfully defined through lip opening. The remaining talker exhibited similar patterns of visual speech production but his/her associated intelligibility was not explained through lip opening. This finding confirms the suggestion that talker variation is the product of many talker-specific factors. Consequently the relative intelligibility of a talker cannot be based upon a single measure alone.

This variation in visible speech production has a direct impact upon the construction of speechreading tests. Bench, Daly, Doyle and Lind (1995) recommended that a group of talkers, with representatives of both genders and a variety of ages, be used when developing a test of speechreading ability. This will ensure that any influence of talker characteristics upon performance will be averaged out across the talkers and will improve the validity of comparing speechreading scores across different tests (Bench et al, 1995). This approach to speechreading testing also improves the ecological validity of the test, as it will represent the ability of the speechreader to understand a similar range of different talkers that they will experience in everyday life. The authors further recommended that these groups of talkers be carefully selected in order to ensure that an average level of speechreadability was maintained (Bench et al, 1995). Essentially, talkers who are extremely difficult to speechread should not be selected. However, in order to ensure that a range of talkers are selected, further information relating to the relative effect of talker characteristics upon visual speech is required so that a more detailed definition of speechreadability can be constructed.

4.1 Experiment 6: A comparison of the effects of talker and accent variation upon speechreading performance

The review of the literature has indicated that talker variation has a significant effect upon both visual speech intelligibility and the construction of speechreading tests. However, the specific factors which might influence intelligibility and the effects of talker variation are undefined. In an effort to further research in this area, Experiment 6 considers the effect of both accent *and* talker variation upon visual speechreading performance. Research within the auditory speech domain has previously examined the effect of increased variation in talker characteristics on talker variability effects. Sommers, Nygaard and Pisoni (1994) examined the impact of speaking rate, talker variation and a combination of talker variation and speech rate on word identification in noise. The multiple talker condition contained utterances from ten talkers, varied on a trial-by-trial basis. The mixed speaking rate was constructed by asking talkers to produce stimuli at three rates – fast, medium and slow. The mixed-rate condition contained a single talker with their speech rate varied on a trial-by-trial basis. Both conditions had a detrimental effect on word identification performance. This finding indicated the associated cognitive cost of compensating for trial-by-trial variation in the acoustic signal (Sommers et al, 1994). Interestingly, performance did not found vary according to speech rate, in that participants were not significantly worse at understanding words spoken at a fast rate compared to a slow rate. Rather, it was the effort of accommodating changes in the speech rate on a trial-by-trial basis that produced the detrimental effect in performance. This highlights variability in the speech signal as opposed to talker-specific characteristics as the main cause of reduced auditory speech intelligibility.

Of specific interest here is the second major finding of the study; that the combination of both factors (speaking rate and talker variability) had an additive effect. Thus, the resulting detriment in performance for the condition in which both speech rate and talker variation were mixed, was larger than for either factor alone. This is illustrated by the bottom panel in Figure 4.3 (Sommers et al, 1994). The results indicate that greater stimulus variability increases the cognitive resource requirements for successful speech processing. Thus, as the number of differences between talkers increase there will be a contingent reduction in accurate speech perception. The effect of talker variation upon speech processing will therefore depend on the number of talker-specific factors that vary within a multiple-talker list. Finally, the additive effect of mixed speech rate and talker variation was maintained across a number of signal-to-noise ratios (i.e. +5 to -5). At the most difficult speech-to-noise ratio of -10 all the individual and additive effects disappeared as speech perception approached floor levels. This indicates that as task difficulty increases any additive effects of talker variability will become less obvious, a potential hazard when applied to speechreading research.

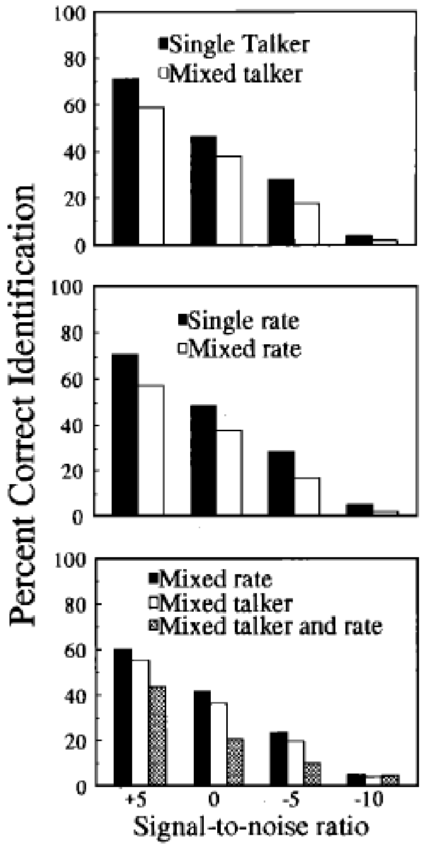


Figure 4.3: (top) Percent correct identification as a function of signal-to-noise ratio for single-talker (solid bars) and mixed-talker (open bars) word lists. (middle) Bars are for single- and mixed-rate word lists. (bottom) Solid and open bars are for mixed-rate and mixed-talker lists respectively. Hatched bars are for conditions in which rate and talker varied simultaneously (Sommers et al, 1994)

The aim of Experiment 6 was to replicate the Yakel et al (2000) study, where speechreading performance was compared across multiple- and single-talker lists, with an additional condition which combines talker and accent variation. It was expected that a combination of talker and accent variation should produce a similar additive effect to that found by combining speech rate and talker variability reported by Sommers and colleagues (1994). Furthermore, due to the influence of accent type upon speechreading performance it was expected that speechreading scores for the ‘mixed accent’ condition would vary significantly on a talker-by-talker basis compared to the ‘single accent’ condition. The experimental hypotheses were:

- Talker variation should reduce speechreading performance in a multiple-talker condition when compared to speechreading performance for a single talker.
- The combination of talker and accent variation should have an additive effect, producing a larger decrement in performance than talker or accent variation alone.
- Speechreading performance should differ according to the accent type of the talker, thus greater variation in performance is expected for the multiple-talker, multiple-accent condition compared to the multiple-talker, single-accent condition.

4.2.1 Method

Subjects

Forty-eight participants were recruited, all were born in England and native English speakers. All subjects reported normal hearing and normal or corrected to normal vision.

Stimuli

Twelve talkers were each recorded articulating 54 BKB sentences (Bench, Kowal & Bamford, 1979). The talkers were split into two groups, the first, ‘same regional accent’ group, consisted of six talkers all with the Nottingham accent. The remaining six talkers formed a ‘mixed regional accent’ group, two talkers had

Glaswegian accents, two had Nottingham accents and two had South of England accents. Each talker's face was fully illuminated using three high power lamps placed at right angles to reduce shadowing. The recordings were made against a neutral white background with only the face and neck of the talker visible. Each recording featured the talker's full face filmed from a camera (Sony Digital Camcorder, DSR-200AP) placed 1.5m away. Each recorded sentence was preceded by 1s of the talker's static face in a closed mouth position, and followed by a further 1s of static footage again in the closed mouth position. Each clip contained only visual information.

For the single-talker, single-accent condition, 12 single-talker sets were produced, one for each talker. In each case the talker was shown articulating all 54 sentences in a random order, but with each sentence shown twice consecutively. This procedure ensured that performance was not biased by selecting only one talker that was shown to all participants. Initially we had envisaged this condition as two separate single-talker conditions, one which would have included talkers with the same (Nottingham) accent type, the other utilising the six talkers with mixed-regional accents. However, it was realised that for each participant the task was effectively the same – a single talker with a single accent. Moreover, individual talkers from the same accent type can still be more or less intelligible depending on individual differences in their articulatory patterns produced within accent variation. Therefore, it was decided to combine the two conditions into the single condition described here.

For the multiple-talker, single-accent condition, one multiple-talker set was produced using the same accent group of six Nottingham talkers. Each talker produced nine of the 54 sentences. The film clips were presented in a random order with the only proviso being that no talker was shown more than twice in a row, ensuring trial-by-trial variation. However, as before each sentence was shown twice consecutively before a new sentence and talker was presented.

For the multiple-talker, multiple accent set, a second multiple-talker set was produced with the mixed accent talker group. Again each talker was shown articulating nine of the 54 sentences, each sentence shown twice. The further proviso, that both talker and accent type must change on a trial-by-trial basis, was included.

Procedure

The paradigm was a between-groups design, with two groups of 12 participants (Groups 2 and 3) and one group of 24 participants (Group 1).

Group 1 was presented with the single-talker, single-accent condition. Each of the 24 participants viewed one talker, meaning that two participants were allocated to each talker. This was to account for individual variation in speechreading ability.

Group 2 was presented with the multiple-talker, single-accent condition comprised of the six Nottingham talkers. Each of the twelve participants viewed all six talkers presented in a random order.

Group 3 viewed the multiple-talker, mixed-accent condition comprised of the two Nottingham talkers, two Glaswegian talkers and two South of England talkers.

All participants were seated at a table three feet in front of an Apple Mac 21 inch computer screen. They were told that they would be viewing a talker's moving face and were instructed to attempt to speechread the sentences which were produced. They could then type their response into the computer keyboard.

4.2.2 Results

Each participant's score was generated using a loose keyword scoring system where errors in morphology are ignored (Bench, Kowal & Bamford, 1979). Each sentence had three keywords, and a point was awarded for each correctly identified keyword, with closely related examples such as plurals also accepted. A participant's score therefore represents the percentage of correctly identified keywords within a sentence set. The total potential number of keywords for each condition was 162.

A mean performance of 28% (45 keywords correct, sd 29) was recorded for the single-talker, single-accent condition, 11% (18 keywords correct, sd 12) for the multiple- talker, single-accent condition and 13% (21 keywords correct, sd 15) for the multiple-talker, mixed-accent condition. These scores indicate a large difference in performance between the multiple- and single-talker conditions, but do not indicate an effect of accent variation. This finding is illustrated by Figure 4.3. Thus, in agreement with Yakel and colleagues (2000), the processing of visual speech was adversely affected by talker variation. However, the influence of accent variation upon the visual signal does not appear to exacerbate this effect.

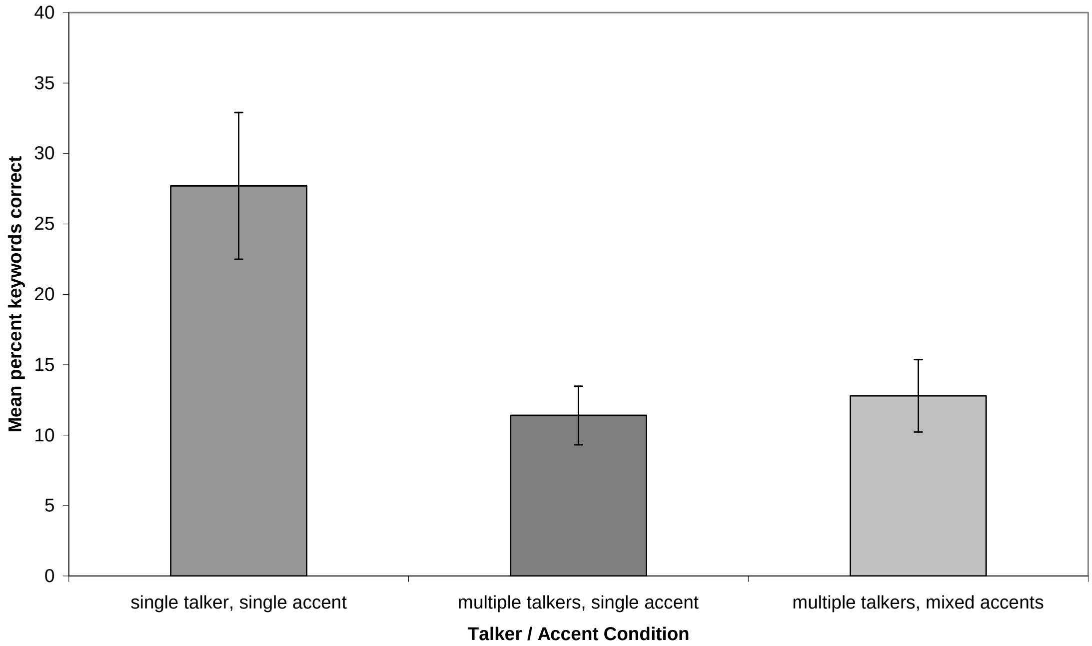


Figure 4.3: Speechreading performance for single- and multiple-talker stimuli lists with single- and mixed-accent variation, shown with standard error bars.

In order to analyse the effect of stimulus condition, a one-way ANOVA was run with the three different conditions as separate levels. The results showed a significant effect of condition: $F(2, 47) = 7.668$, $p < 0.01$. Post-hoc analysis using the Tukey HSD procedure revealed two significant pairwise effects, as illustrated by Table 4.1.

(1) Condition	(2) Condition	Mean Difference	p-value
Single talker, single accent	Multiple talkers, single accent Multiple talkers, mixed accents	26.46* 24.21*	.005 .011
Multiple talkers, single accent	Multiple talkers, mixed accents	2.25	.97

Table 4.1: Results of Tukey pairwise comparisons of means for each subject group on a talker and accent variation basis

The results show a very clear detrimental effect of talker variation on speechreading performance, with the significantly more accurate performance recorded for the single-talker condition than either of the multiple-talker conditions. This effect was not altered by the introduction of accent variation since there was no significant difference in performance levels between the single- and mixed-accent multiple-talker conditions. Thus, in comparison to auditory research on the subject, there does not appear to be any additive effect from two sources of variation in the visual speech signal.

The research detailed within Chapter 3 led to the expectation that accent type would significantly influence speechreading performance in the multiple-accent condition. To investigate the influence of accent type, speechreading performance was plotted across talkers for the two multiple-talker conditions, as illustrated by Figure 4.4. The Figure shows clear differences in speechreading performance across the two conditions. This replicates the result found in Chapter 3, Experiment 5, where variation was found between talkers both within and between the accent groups. This illustrates the influence of indexical information other than accent type on a talker's relative speechreadability.

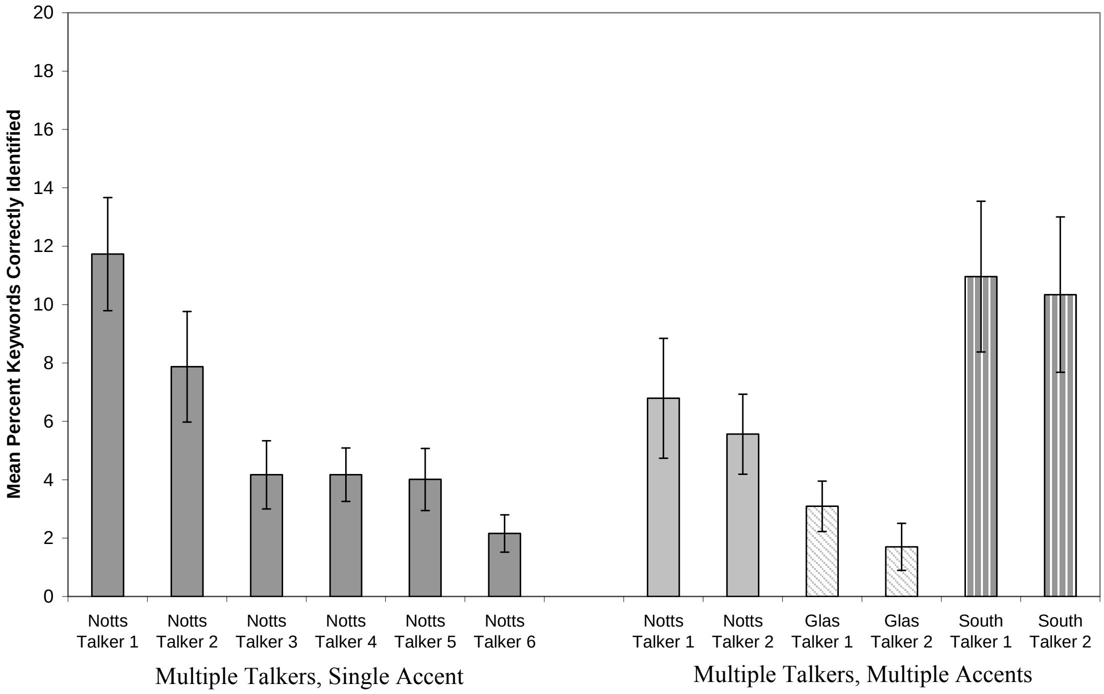


Figure 4.4: Speechreading performance plotted on a talker-by-talker basis for two conditions, multiple-talkers, single-accent and multiple-talkers, multiple-accents, shown with standard error bars.

To evaluate whether the differences between talkers were significant, two separate one-way ANOVA (6 talkers) were run, one for each condition. The results showed a main effect for speechreading performance across the talkers, both for the multiple-talker, single-accent condition: $F(5, 55) = 10.813$, $p < 0.05$. and for the multiple-talker, mixed-accent condition: $F(5, 55) = 6.253$, $p < 0.05$. Post-hoc testing using the LSD method, revealed the following: In the multiple-talker, single-accent condition, performance for Nottingham Talker 1 was significantly better than the performance for Nottingham Talkers 3 – 6 (see Appendix 5 for full table of results). Performance for Nottingham Talker 2 was also significantly better than for Nottingham Talkers 4 – 6. In the multiple-talker, mixed-accent condition, the talkers from each accent group were paired together.

Scores for Nottingham Talkers 1 and 2 were not different from one another, but were significantly better than those for the Glaswegian talkers and significantly worse than the South of England Talkers. The same was found for the two Glaswegian and South of England Talkers i.e. the talkers from the same regional accent type were more similar to one another than the talkers from a different accent type. This suggests that the accent type of the talkers influenced their visual speech intelligibility.

Finally, the possibility was considered that there may have been a sentence-by-talker interaction, such that talkers could have been easier to speechread simply because they uttered sentences which were visually easy to identify. However, consideration of the stimuli showed that the set of sentences selected for Nottingham Talkers 1 and 2, and the South of England Talkers 1 and 2 were different, disputing that suggestion.

4.2.3 Discussion

The results confirm previous research (Yakel et al, 2000) that has shown an effect of talker variation upon speechreading performance. The detrimental influence of talker variability indicates, similar to findings within auditory research (Mullenix, Pisoni & Martin, 1989), that the visual speech processing system is adversely affected by consistent talker-by-talker variations in the visual signal. Yakel and colleagues (2000) argued that there were two explanations for this effect. The first suggests that the detrimental influence of multiple-talker lists is due to the encoding of talker indexical information which requires the utilisation of cognitive resources. Essentially, as previously stated by Mullenix et al (1989) the encoding of talker characteristics utilises memory and cognitive resources, reducing the resources available for speech processing. Moreover, the encoded talker information could then interfere with the subsequent linguistic analysis of later talkers. The present results would appear to uphold this hypothesis. However, the second explanation for the results, namely that talker familiarisation facilitated speechreading performance, as opposed to talker variation inhibiting performance, is equally upheld by the present results. Although the results of Experiment 6 do not distinguish between the two potential explanations given for the performance differences between single- and multiple-talker conditions, they do provide further evidence for the importance of talker characteristics in visual speech perception.

The second issue of interest was the influence of the combined effects of talker and accent variation upon speechreading performance. Surprisingly, there was no additive effect of the two forms of variation. Both multiple-talker conditions produced similar levels of performance, irrespective of whether accent type was varied or not. This disagrees with the result of the Sommers et al (1994) study using auditory speech. There are three alternative explanations for this null result. First, it is possible that accent type was not automatically processed by the visual system, allowing the perceiver to ignore accent differences in order to perceive the linguistic content of the message. This explanation is unlikely based on the results of Chapters 2 and 3, where significant effects of accent type on visual speech intelligibility were found. However, if we assume that accent type forms an integral part of the visual signal the alternative explanation for the null result is that accent type has a greater influence upon speechreading performance than accent variation.

Essentially, a talker's accent type will influence the intelligibility of their visual speech. However, when an observer is presented with multiple accent types the individual impact of each accent type is not changed by this variation, rather each talker is still influenced by their accent type but the change in accent type from talker to talker does not increase its effect. This second suggestion is supported by research conducted on the effects of increased fundamental frequency or rescaling formant frequencies, or a combination of both, on word identification (Diehl, Souther & Convis, 1980). Diehl and colleagues (1980) found no additive effect of combining formant and fundamental frequency differences upon word identification, instead they reported a slight *reduction* in multiple-talker effects when stimulus variability was increased. The authors suggested that the lack of additive effect was due to the listener selectively disregarding previous talker-specific information and reinitiating the encoding process for each change in a talker characteristic. Thus, the reduction in performance caused by multiple talkers is essentially derived from the listener identifying a change in talker and beginning the normalisation or encoding process anew. The number of levels along which the stimulus changes should have no effect as in each case the listener disregards the previous talker-specific information (Diehl et al., 1980). It is possible that a similar effect occurred within Experiment 6 producing the recorded null result for additive effects.

Finally, the third possible explanation lies in the poor performance recorded for both multiple-talker groups. It was noted that within the Sommers et al (1994) study, the additive effect of speech rate and talker variability was reduced when overall performance was lowered through a low (-10) signal-to-noise ratio, with the reduction attributed to floor effects. Potentially, the additive effect of talker and accent variation in Experiment 6 could have been reduced through similar floor effects (performance was only 11-13% correct). Lander and

Davies (2008) avoided such floor effects by combining their visual stimulus with a low auditory signal (~20dB at 20m). Further research in this area utilising a similar methodology, or improving speechreading performance in some other way (i.e. context, see Chapter 5, Experiment 8) may find an additive effect once performance has been improved away from floor level.

The finding that speechreading performance across talkers in the multiple-talker, mixed-accent group varied according to accent type suggested that although accent variation did not have a direct effect, a talker's accent type still influenced visual speech intelligibility. This consolidates the argument put forward in Chapter 3, Experiment 5, that there is an inter-relationship between accent type and visual speech intelligibility. However, the effect of accent type does not appear to extend into an additive effect on the cognitive load of processing multiple sets of talker information. This result differs from auditory research on the topic (Sommers et al, 1994) and indicates a further potential dissociation between auditory and visual processing. Sommers and colleagues (1994) reported that speech rate did not directly influence acoustic processing, but that variation in speech rate across trials had a detrimental effect. The results of the present experiment are the opposite, with accent type directly influencing intelligibility but accent variation producing no additive effect above talker variation.

It is possible that the accent types used within the multiple-accent condition were insufficiently different to elicit an additive effect with talker variation. It is possible that the use of an accent from one of the remaining main accent regions, such as the Welsh or Irish accent, which differ along a greater number of dimensions to the Nottingham accent than the South of England accent used here, would have produced a larger effect.

Finally, an interesting point to note is that the results of Experiment 6 offer an explanation for the low performance scores reported in Experiments 4 and 5. The utilisation of 12 talkers evidently had a detrimental effect on observer's speechreading ability in both experiments, reducing performance levels. This finding should be considered in the development of future speechreading tests, where speechreading performance will be reduced if more than one talker utters the speech stimuli.

4.3 General Summary

To conclude, the results presented in Chapter 4 illustrate the complexity of the relationship between observer and talker in visual speech perception. It is evident that talker variability has a detrimental effect on speechreading performance. However, the relative influence of various indexical properties upon talker variability remains unclear. Moreover, it is apparent that the responses of the visual and auditory processing systems to variability differ. Whereas a combination of talker characteristics exert an additive effect upon the auditory processing system, it appears that although characteristics such as accent type influence the visual intelligibility of talkers, these factors do not exert a direct effect on talker variability effects.

Chapter 5. The Impact of Increased Exposure, Contextual Cues and Repetition upon Accent Effects in Visual Speech

'Mans mind, once stretched by a new idea, never regains its original dimensions'.

Oliver Wendell Holmes (1809 – 1894)

Despite the variability inherent in the audiovisual speech signal, the majority of observers are able to compensate for talker variation in order to comprehend a talker's message. This ability was originally thought to be part of the process of speech normalisation, where the observer filters idiosyncratic talker characteristics from the speech signal in order to process speech in an abstract form (Nygaard & Pisoni, 1998). However, more recent research suggests that rather than removing variability from the speech signal, the speech perception system instead directly encodes talker indexical information, using that information to inform perception and improve speech intelligibility (Pisoni, 1997).

An extension of this theory was put forward by Kraljic and Samuel (2005). They suggested that the internal representations of speech remain flexible and can use new information, such as a talker's individual mode of speech production, to adjust to new talkers and situations. Essentially, they propose that the perceptual system is able to incorporate new tokens of speech and modify internal speech prototypes, using the information to adjust their parameters of accepted speech and improve comprehension. An example of such dynamic adjustment was reported in their recent study, which found experience with a particular talker could alter the boundary between different phonetic categories (Kraljic & Samuel, 2005). The study exposed listeners to a single talker producing /s/ or /ʃ/ in such a way that both sounds were ambiguous and then tested their categorisation of sounds along a /s/ - /ʃ/ continuum, produced by the same talker. The results showed that the participant's perception of the sound was altered by their exposure to the ambiguous phoneme, with a larger proportion of ambiguous sounds being categorised as /s/ or /ʃ/ depending on the participant's previous exposure (Kraljic & Samuel, 2005). Furthermore, further exposure to the same or different talker producing unambiguous /s/ or /ʃ/ sounds did not alter the participants' newly formed ambiguous categorisations. The authors suggest that an observer's internal representations of speech sounds adjust according to a talker's output and are stored on a long-term basis, consequently accessed whenever the same output is encountered. They argue that such an adjustment indicates that active perceptual learning is taking place, as opposed to passive adaptation.

Further research in this area indicates that the parameters of an unfamiliar accent type can also be encoded within the speech processing system to aid comprehension. Clarke and Garrett (2004) examined the reaction times of native English speakers to non-native renditions of English (Spanish speakers producing English sentences). Performance was measured by recording a participant's responses to visual probes (text words appearing on screen) the task being to identify the visual probe that matched the last word of a preceding spoken sentence. The results showed that participants were significantly faster to respond to sentences produced in a native as opposed to a non-native accent (see Chapter 2, Introduction for a full description). However, after approximately one minute of exposure to accented speech, participants showed a significant improvement in performance with reaction times recorded as at the same speed as responses to native speech. The study illustrates the adaptability of the auditory system, with learning and consequential alteration of phonetic categories occurring within one minute of exposure to non-native accented speech (Clarke & Garrett, 2004).

Another study examined lexical adaptation to computer-generated accent changes (Maye, Aslin & Tanenhaus, 2008). For example, one such generated accent was the lowered front vowel accent where ‘witch’ becomes ‘wetch’ etc. Participants were presented with 20 minutes of speech presented in their ‘home’ or familiar regional accent, followed by an auditory lexical decision task (word / non-word classification). They were then exposed to 20 minutes of the computer modified speech and asked to complete the same lexical decision task. The results showed that participants changed their classification of phonetic forms after exposure to the computer modified accent, leading to increased classifications of words as opposed to non-words, for example, ‘wetch’ would be classified as a non-word in the first session but as a word in the second session (Maye, Aslin & Tanenhaus, 2008, see also; Bardhun, Aslin & Tanenhaus, 2006). The authors suggest that listeners are able to adjust their representations of phonetic forms in response to a new accent type. Moreover, this alteration is specific to the accent that is encountered, rather than simply an acceptance of an ‘odd’ pronunciation or relaxation of phonetic boundaries. This interpretation is supported by research examining listeners categorisation of vowel sounds, which found that participants adjusted their decisions based on the accent of the sentence in which the vowel was imbedded (Evans & Iverson, 2003).

The reviewed literature indicates that the auditory speech perception system is highly adaptable, both to new talkers and to new accent types, with less than one minute of exposure required in some cases (Clarke & Garrett, 2004) for learning to take place. There has been no prior research on the ability of the visual system to encode accent related variations in the visual signal. However, due to the close integration of the visual and auditory systems, it is likely that a similar pattern of learning will be found. Support for this suggestion is found within the visual literature, which indicates that talker familiarity can improve perception of visual speech. For example, Yakel et al (2000) suggest familiarity with a talker as a potential explanation for their finding that speechreading performance for a single talker is significantly better than recorded performance using multiple talkers. Effects of talker familiarity can also found within research that has shown speeded classification of visually presented vowels to be facilitated by familiarity with the talker (Schweinberger & Soukup, 1998).

A recent study also found an advantage of talker familiarity for speechreading words (Lander & Davis, 2008). The study consisted of five sections, these were split into three test sessions containing word lists, and two 'training' sessions where an audiovisual presentation of a two minute story was produced by a single talker (see Chapter 3, Experiment 5 for a full description). The results showed that speechreading accuracy improved to a greater extent when the same talker was used for all five sections (mean improvement = 18%) than when a different talker was used in the training sessions from the test sessions (improvement = 6%). The authors suggest that increased exposure to a talker allows observers to adapt to the speaking style of that individual, hence improving their ability to speechread that talker (Lander & Davies, 2008).

All of the described studies indicate that continued exposure to a talker leads to familiarity with their mode of speech production, with a consequential advantage in visual speech perception. The aim of the research presented in

Chapter 5 was to determine if exposure to an unfamiliar and difficult (Glaswegian) accent type would improve speechreading performance for that accent, with consideration given to the various methods for achieving this improvement.

5.1 Experiment 7: The effect of prolonged exposure on accent related effects in speechreading performance

In the auditory domain exposure alone, without any feedback or other training mechanisms, is enough to improve a listener's ability to comprehend a talker. To clarify, exposure, as I discuss it here, relates to prolonged exposure to a stimulus whilst completing a task, as such, it is a form of practice or training. Other training mechanisms include offering feedback after each stimulus and repeating a task unit until the participant has successfully completed it. However, the first experiment detailed here is an investigation into the effects of exposure alone upon an observer's ability to comprehend visual speech produced in an unfamiliar and previously identified as difficult Glaswegian accent.

Three types of exposure were used within the experiment with speech presented in the visual, auditory and audio-visual modality. Thus, participants were tested using visual speech, then familiarised with the accent type using visual, auditory or audiovisual stimuli. The reason for this was three-fold. First, experience speechreading a talker was shown to facilitate subsequent comprehension of speech in noise produced by the same talker (Rosenblum, Miller & Sanchez, 2007). Second, research indicates that information specific to a talker is available across both modalities, allowing cross-modal talker identification (identifying a talker's face from their voice and vice-versa) (Kamachi, Hill, Lander & Vatikiotis-Bateson, 2003). Third, work on the recalibration of perception based on mis-matched visual and auditory signals (Bertelson, Vroomen & Gelder, 2003) has shown that the perception of auditory speech can be adjusted through exposure to an altered visual signal. It is possible, for the purposes of the present experiment, that the opposite could also be true; with exposure to the auditory signal altering perception of the visual modality. Specifically, knowledge of the auditory component of an unfamiliar accent may improve speechreading performance for that accent.

The experimental hypotheses were:

- An extended period of exposure to the Glaswegian accent should result in some improvement in speechreading accuracy, produced through the observers increased familiarity with the talkers' mode of visual speech production.
- Exposure to talkers using audiovisual or auditory stimulus should result in a similar improvement in speechreading accuracy to that found through using visual speech.

5.1.1 Method

Participants

Thirty participants were recruited for the study, all were native English speakers from Nottingham and reported normal hearing and normal (or corrected to normal) vision.

Stimuli

In total, 308 BKB (Bench, Kowal & Bamford, 1979) sentences were recorded using eight talkers (as described in Chapter 3), four of whom had a Glaswegian (Central Scottish) accent, the remaining four having a Nottingham

(East-Midlands) accent. Each talker's face was fully illuminated and recorded against a light background with only the face and neck visible for each sentence recording. Each sentence was recorded using a microphone, with the sound later separated from the visual element of the stimulus in order to allow utilisation of both. The pre, mid and post tests of speechreading ability constituted 168 sentences (56 sentences for each test, every talker producing seven sentences within each set).

Each sentence set was balanced across talkers by matching sentences for phonetic content. For example, if Nottingham talker 1 produced the sentence ‘the boy forgot his book’, Glaswegian talker 1 would produce the sentence ‘a boy broke the fence’. Thus sentences were matched for similar visual productions i.e. the bilabial /b/ was present in both sentences twice, in ‘boy’ on both occasions and then as ‘broke’ and as ‘book’ alternatively. These matches were made on the basis of not only phonetic similarity, but on previously identified levels of sentence difficulty (MacLeod & Summerfield, 1987). Thus, within each sentence set there were identical numbers of ‘difficult’ and ‘easy’ sentences, chosen according to the levels set out by MacLeod and Summerfield (1987). To further counterbalance any remaining effects due to item difficulty, the order in which the test sessions were presented to each participant was randomised. Hence participant 1 might view test sessions 1, 2 and 3 in that order, then participant 2 would view the test sessions in the order 3, 1, 2 and so on.

The remaining 140 BKB sentences were recorded for use as two training sessions, produced by the four Glaswegian talkers (same set of sentences used in both training sessions, but in the second training session the talker producing each sentence was different).

Procedure

The experiment was made up of five parts: pre-test, training, mid-test, training and post-test. The test sentences were all presented in the visual modality, whereas the training sentences were presented in the visual, audiovisual or auditory modality depending on the participant's group designation. Each participant was seated at a table directly in front of the view screen and instructed to watch, or listen, to each clip carefully. The task for both test and training sessions was the same, in each case participants were asked to identify what the talker had said and then type their response on the computer keyboard. No feedback was given at any stage. The experimental procedure is illustrated by Figure 5.1.

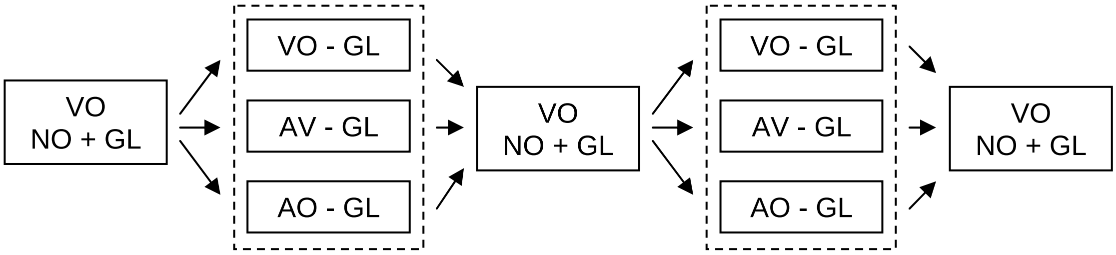


Figure 5.1: Order of presentation of stimuli to participants (VO: visual only, AO: auditory only, AV: audiovisual, NO: Nottingham accent, GL: Glasgow accent)

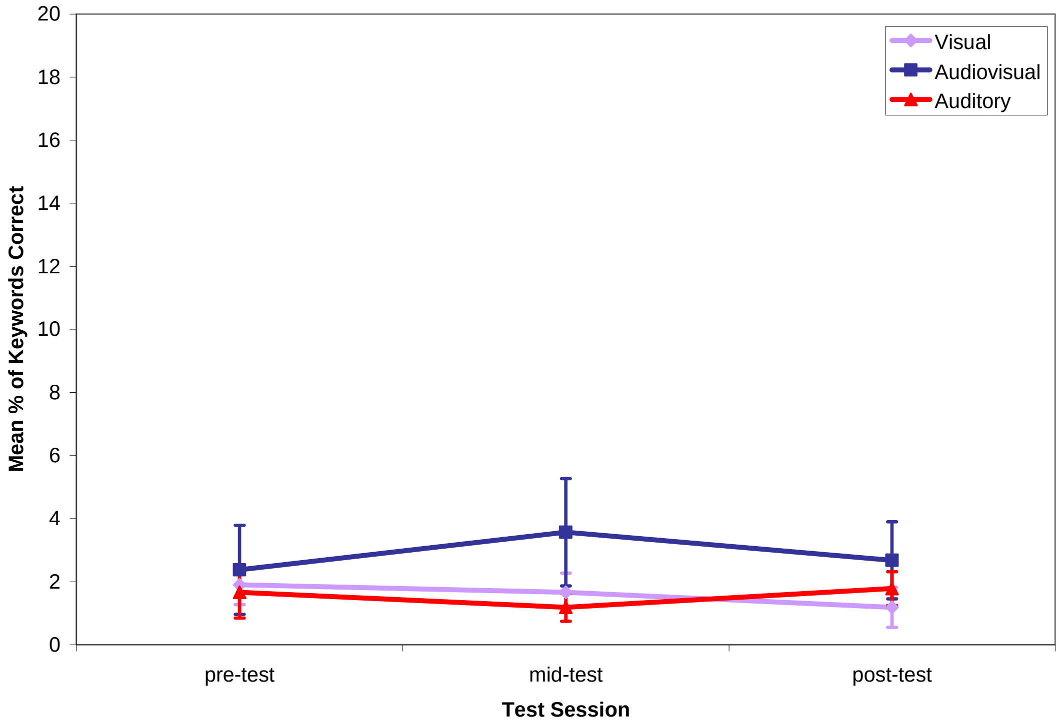
5.1.2 Results

Each participant's score was generated using a loose keyword scoring system with errors in morphology ignored with closely related examples such as plurals accepted. A point was awarded for each correctly identified keyword. A participant's score therefore represents the number of correctly identified keywords within a sentence set, in this case the potential total of keywords for each test set was 168.

The results were analysed using a three-factor mixed factorial (3 training modality x 2 accent type x 3 testing session) ANOVA. The mean scores for the three testing sessions were not found to differ at a significant level: $F(2, 54) = .105$; $p > 0.05$. Thus the scores for all three groups (pre test mean: 3.3, mid test mean: 3.6, post test mean: 3.2 keywords correct) did not improve significantly through training. There was a significant effect of accent type upon the results: $F(1, 27) = 25.036$; $p < 0.05$, indicating that participants keyword scores for the Glaswegian talkers (mean: 3.37) were significantly poorer than their scores for the Nottingham talkers (mean: 6.12). There was no significant effect of training type: $F(1, 27) = 2.025$; $p > 0.05$. Finally, no significant interactions were found between the three factors. These results are illustrated by Figure 5.2.

An important consideration noted whilst analysing this data was that the keyword accuracy scores were very low, with a range of only 4 – 10 keywords correct out of a possible 168 across both accent groups, representing an average performance level of 2 – 6% correct. Therefore it is highly likely that the results are influenced by floor effects.

GLASWEGIAN TALKERS



NOTTINGHAM TALKERS

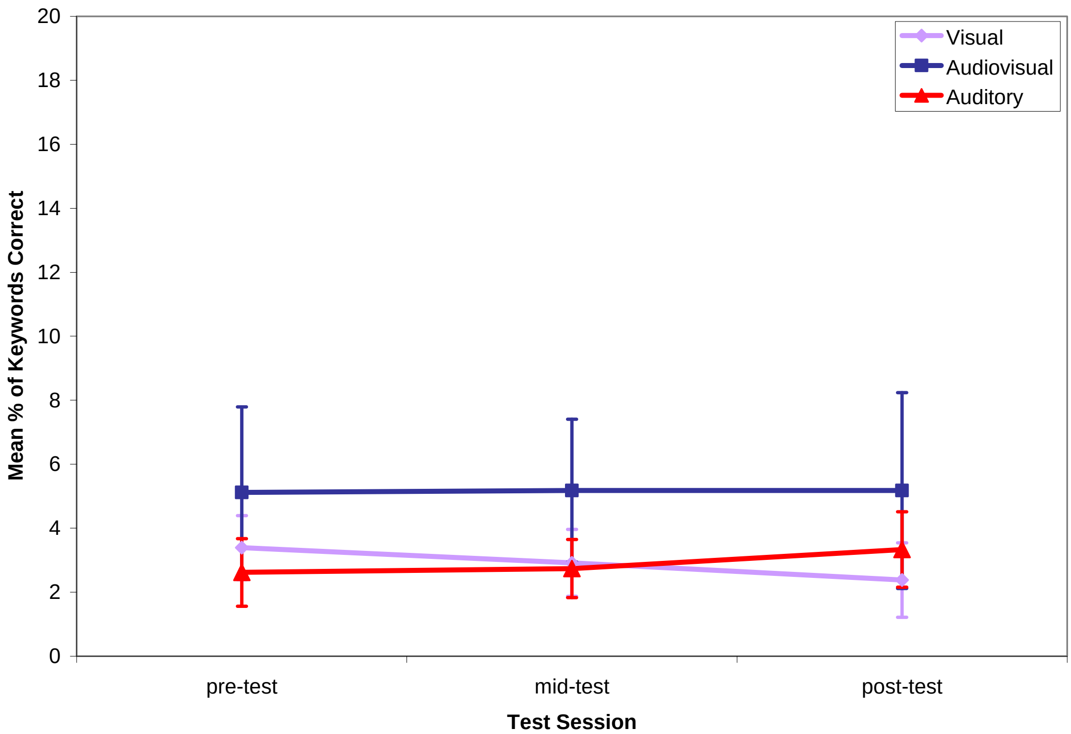


Figure 5.2: Mean percentage of keywords correctly identified for Glaswegian and Nottingham talkers across three test sessions, with standard error bars

5.1.3 Discussion

These results suggest that exposure alone does not appear to be enough to mediate accent effects in visual speech. Certainly performance was not found to differ for the four Glaswegian talkers across the three testing sessions and the difference in performance levels between the Glaswegian and Nottingham talkers did not alter across test sessions. Therefore we are unable to reject the null hypothesis in this case. However it should be noted that, due to the low speechreading performance of the participants, floor effects may be responsible for the non-significant results. The possibility that exposure might have an influence when such effects are not present should, therefore, be considered.

Previous auditory research posited that adaptation to accent occurs over three stages; initial disruption of performance, adaptation through the encoding of specific accent information and eventual recovery of performance (Floccia et al, 2006). This recovery of performance has been found to occur within studies utilising both lexical decision (Floccia et al, 2006) and word identification (Clarke & Garrett, 2004) tasks. The null result reported here could indicate that the visual processing system is less able to adapt to accent change than its auditory counterpart. However, the nature of the task utilised in the present study differs from the requirements of the two studies noted above. First, the stimuli presented were of sentence length, requiring protracted processing, the results of which function as a measure of visual speech intelligibility. In comparison, the studies detailed above represent a measure of comprehensibility, or how difficult speech is to process, usually measured by the length of time required for that processing to occur. It is possible that there is a level of dissociation between the two, with intelligibility representing the more difficult facet of speech perception.

Second, many of the auditory studies on this topic (Kraljic & Samuel, 2005; Bardhun, Aslin & Tanenhaus, 2005) have utilised variations in the speech of a single talker in their research. Coupled with the research on visual speech which indicates that a single talker is easier to speechread than a group of varied multiple talkers (Yakel et al, 2000), it is possible that the lack of a training effect in the present study was caused by the utilisation of multiple talkers. Potentially the cognitive load of processing and learning the unfamiliar speaking styles of four talkers was too much for the visual processing system in this case, leading to the negative result. However, it should be remembered that the aim of this experiment was to examine the nature of accent familiarity. In order to do this it was necessary to utilise multiple talkers to ensure that the familiarity being measured was related to accent type as opposed to single talker familiarity. Furthermore, theories of speech perception indicate that once the parameters of an unfamiliar accent have been learned, the benefit of the altered speech representations should generalise across all talkers with that accent (Norris, McQueen & Cutler, 2003, see also, Gass & Varonis, 1984). Therefore, the use of multiple talkers cannot completely explain the null result.

Third, the low speechreading scores recorded may have had an influence upon results; each of the talkers was chosen on the basis of their accent type rather than their speechreadability (as discussed in Chapter 3, Experiment 5), leading to low speechreading scores both in the present and previous studies within this thesis. It is possible that the low level of information being processed from the visual signal meant that the task was too difficult to allow learning to take place. The question of exposure to accent types when the general level of speechreading performance is better will be discussed in the next few studies within this chapter.

Finally, the modality examined here differs from the auditory signal in several ways; the limits of the visual signal, with regards to the lack of visible articulation of certain aspects of speech (Summerfield, 1991) for example. The level of information carried within the visual signal relating to accent type may be insufficient to allow adaptation to occur within the limited time frame used here. Long term exposure to an accent type over several days may be more successful at bringing about improvement. Certainly past research investigating speechreading training has usually lasted over a period of days (Walden, Erdman, Montgomery, Schwartz & Prosek, 1981), therefore lack of exposure time may have contributed to the negative result in this case.

To summarise: The results indicate that an unfamiliar accent type has a negative effect upon speechreading performance, one which cannot be mediated through exposure alone. This brings us to our next study as, although it is possible that accent is a greater disadvantage for speechreaders than for listeners utilising the auditory component of speech, it is likely that those same speechreaders will attempt to improve their comprehension of the message through various strategies that are available when they engage in everyday conversation. The first of these strategies involves the use of contextual cues, something that will be discussed in the next two experiments within this chapter.

5.2 Experiment 8: The impact of contextual constraints upon regional accent effects in speechreading performance

The results of the previous experiment indicate that exposure alone may not be sufficient to improve perception of an unfamiliar accent in visual speech. As mentioned, it is possible that this is partly due to the low performance levels. Essentially, if the participants were unable to extract any useful information from the visual signal to begin with, further exposure without any kind of training mechanism will do little to alleviate accent effects. The next experiment in this chapter represents an attempt to reduce some of the inherent task difficulty through the utilisation of context, in this case a cue word presented before the visual stimulus. The word represents a topical constraint (Boothroyd, 1988), providing the observer with a subject area that the presented visual sentence will fall within. For example, the word presented might be ‘pets’ and the visual sentence afterward ‘the puppy chased the ball’. This type of context acts to diminish uncertainty about message content by reducing the possibilities before the visual stimulus is presented (Boothroyd, 1988), thereby providing the observer with a ‘clue’ as to the nature of the sentence. By presenting participants with this small level of context we hope to improve task performance, determine whether contextual cues have any influence upon accent effects and, finally, determine if exposure to an accent type is more effective when initial performance levels have improved.

A large body of research exists on the advantages of context in auditory speech perception. For example, context has been utilised as an aid for hearing-impaired participants, with sentences (Most & Adi-Bensaid, 2001; Grant & Seitz, 2000) or preceding questions (Flynn & Dowell, 1999) improving word comprehension when the auditory signal is degraded. It is likely that such augmentation comes about through linguistic redundancy, which relates to the use of language constraints to compensate for any gaps in perception caused by difficulty in perceiving the acoustic speech signal (Boothroyd, 1988). Essentially, context helps an observer form expectations about the content of a message and then use those expectations to aid comprehension.

Contextual cues can also be used to aid detection of mispronunciations in auditory speech and to improve the segmentation of speech where mispronunciation makes speech ambiguous (Cole, Jakimik & Cooper, 1980). Thus mispronounced speech units that could be perceived as one or two words were presented within the context of a story to enable participants to decipher their meaning. For example, the non-word *'carko'* could be perceived as *'car go'* or *'cargo'* depending on whether it was presented within a story about a car about to drive on a ferry, or about a shipment of cargo (Cole et al, 1980). Context speeded recognition of the mispronunciation and aided the participant in prescribing meaning to the non-word, illustrating the use of higher order information (contextual constraints) in the perception of continuous speech.

Contextual effects have also been studied using brain potentials, usually utilising a manipulation of the 'cloze procedure' (Taylor, 1953). The original procedure involved the removal of words from a sentence or passage of text. The task was to 'fill in the blanks' using contextual cues provided by the surrounding text (Taylor, 1953). A variety of manipulations of this method have been conducted, using visually presented text (Kutas & Hillyard, 1980) auditory speech (Connolly & Phillips, 1994) and visual speech (Calabresi, 2007). Most often, context is provided through a preceding sentence, followed by a semantically congruous (high cloze probability) or incongruous word (low cloze probability). For example, the sentence *'I take coffee with cream and…'* could be followed by the congruent *'sugar'* or the incongruent *'machine'* (Kutas & Hillyard, 1980). Research is this area has found that certain areas of the scalp, recorded through event-related brain potentials (ERP's) are sensitive to contextually generated expectations as detailed above (Kutas & Hillyard, 1980), with researchers showing a particular interest in the generation of a negative brain wave (N400) in response to semantic incongruence.

The N400 has been recorded in response to a semantically incongruent final word of a sentence, presented both as text (Kutas & Hillyard, 1980) and when spoken (auditory stimulus) (Connolly & Phillips, 1994). Interestingly, the N400 relates only to semantic incongruence; a final word that has an incongruent initial phoneme (not the expected high cloze phoneme) but is semantically acceptable (i.e. *'Don caught the ball with his* ***glove****',* Connolly & Phillips, 1994) does not elicit the N400, indicating that it is a specific reaction to semantic content as opposed to a general reaction to an unexpected final word. Essentially, the N400 appears to reflect interrupted speech processing caused by the semantically inappropriate word, necessitating increased processing as the individual attempts to make sense of the non-word (Kutas & Hillyard, 1980).

In comparison, the work investigating the N400 using visual speech (Calabresi, 2007) did not find the expected response, with no N400 elicited using visual semantic incongruence. The author suggests that although the visual and auditory modalities contain equivalent stages of processing, the relative weight of each of these stages differ. Essentially, difficulty in accurately perceiving visual speech means that although contextual information is used, it is not integrated at an early stage of processing. Rather, several possibilities regarding the stimulus are held in memory and then reanalysed at a later stage together with the contextual information, then a decision about the content of the stimulus is reached (Calabresi, 2007). Thus contextual information is utilised at a later stage than within auditory processing eliciting a recorded P600 response, indicating a period of reanalysis of the stimulus, rather than the N400 response recorded within auditory literature.

Providing context has also been used to alleviate comprehension difficulties caused by an unfamiliar auditory accent. For example, Labov (1989) designed a gating experiment to determine if the addition of contextual information would reduce the effect of accent upon word comprehension (Labov, 1989; Labov & Ash, 1997). The study utilised talkers exhibiting Southern Shift (characteristics of a Southern American accent), presented to participants unfamiliar with the accent. This allowed the authors to document the effect of accent change, namely differences in vowel production, upon comprehension. The results showed that, although increasing levels of context (phrase, then sentence) improved the intelligibility of speech produced in an unfamiliar accent, context alone was not enough to compensate for those effects completely. Indeed, once a particular vowel sound had been incorrectly identified by a subject unfamiliar with the accent, they seemed averse to altering their perception of it, even when the preceding phrase made their assessment of the sound meaningless. Thus, the acoustic properties of speech still influence perception when the expected meaning of a word is constrained by context.

The effect of context on visual speech comprehension has also been investigated, with similar results to those found in auditory speech. For example, a question-answer format with an initiating utterance (question) providing context, was found to improve speechreading performance for visual sentences (Erber, 1992). Moreover, research indicates that context in the form of a priming word (presented as text) can lead to improved perception of spoken visual words (Lansing & Helgeson, 1995). However, factors such as the level of visibility of a presented word (i.e. how easy that word is to speechread) can also have an effect. This was illustrated when neutral context (no useful semantic information provided by the cue) was utilized. In that circumstance the relative visibility of the words used had the most impact upon recognition levels; indicating that in the absence of any external information the observer relied upon sensory factors relating to the visible articulatory movements. This suggests that the relative intelligibility of a talker, as measured by their visible articulations, might also influence contextual effects. The authors concluded that both sensory and linguistic (in the form of context) factors have an effect on an individual's ability to speechread (Lansing & Helgeson, 1995).

Finally, research has also indicated that a combination of cue words (i.e. the sentence 'the tie is black' might be cued by the words 'clothes' and 'colour') (Lidestam, Lyxell & Lundeberg, 2001) was enough to improve speechreading performance for sentences. The authors also note that the effect of context appeared to vary according to task difficulty; the more difficult the task the greater the improvement provided by contextual cues (Lidestam, Lyxell & Lundeberg, 2001). This has ramifications for the present study as the talkers with a Glaswegian accent represent a more difficult speechreading task in this case, when compared to those talkers with a Nottingham accent. Thus, contextual cues may aid perception of talkers with a Glaswegian accent to a greater degree than talkers with a Nottingham accent, potentially reducing accent effects on performance. Alternatively, if the results of auditory research are replicated (Labov, 1989), the constraints provided by context may be insufficient to reduce accent effects. With this in mind, the study presented here was designed to purely investigate the effects of contextual cues upon speechreading performance across two accent types (Glaswegian and Nottingham) without the element of training. This was to allow us to determine the effect, if any, of context upon accent effects in visual speech. The experimental hypotheses were as follows:

- Contextual cues should improve general speechreading performance
- Performance should increase to a greater extent for the Glaswegian talkers in comparison to the Nottingham talkers, reducing the effect of regional accent
- The effect of context should differ across talkers due to interactions with other factors of talker variability (visible articulation)

5.2.1 Method

Participants

Fifteen participants were recruited for the study, all were native English speakers and reported good hearing and normal (or corrected to normal) vision.

Stimuli

The same sets of sentences used in Experiment 5 were used again here, with both the talkers and the sentence contents being identical. Thus, 240 sentences from the BKB set (Bench, Kowal & Bamford, 1979) were recorded, with each talker's face fully illuminated and recorded against a light background with only the face and neck visible. The sentences were split into twelve lists comprising 20 sentences each, every list spoken by a different talker. Each sentence was preceded by a cue word that provided topical constraint for the sentence that appeared afterwards, e.g. the word ‘property’ might precede the sentence ‘the family bought a house’.

The talkers used for the recordings were grouped as follows: six talkers (three male, three female) with Nottingham accents, while the remaining six (four male and two females) had Glaswegian accents, all were talkers used in both the previous experiment and those described in Chapter 3, Experiments 4 and 5. In order to counterbalance the sentences across accent type, two lists were produced – list A and list B, each containing 120 sentences. Seven of the participants viewed list A produced by the six Glaswegian talkers and list B produced by the six Nottingham talkers. The remaining eight participants viewed the opposite, with list A produced by the six Nottingham talkers and list B produced by the six Glaswegian talkers.

Each sentence was preceded by 1s of the talker's static face, and followed by a further 1s of static recording. Each clip contained only visual information.

Procedure

Participants were seated at a table directly in front of the computer screen, and instructed that they would first be shown a 'cue word', described as providing a 'clue' as to the nature of the sentence, after the cue word had appeared a video clip would follow shortly after, with each talker producing one sentence per video clip, which they were asked to watch carefully. Their task was to attempt to identify what the speaker had said and type their response, as illustrated in Figure 5.3. They were not required to understand the entire sentence; any word that was typed in was recorded. All fifteen subjects viewed the clips presented in a randomised order.

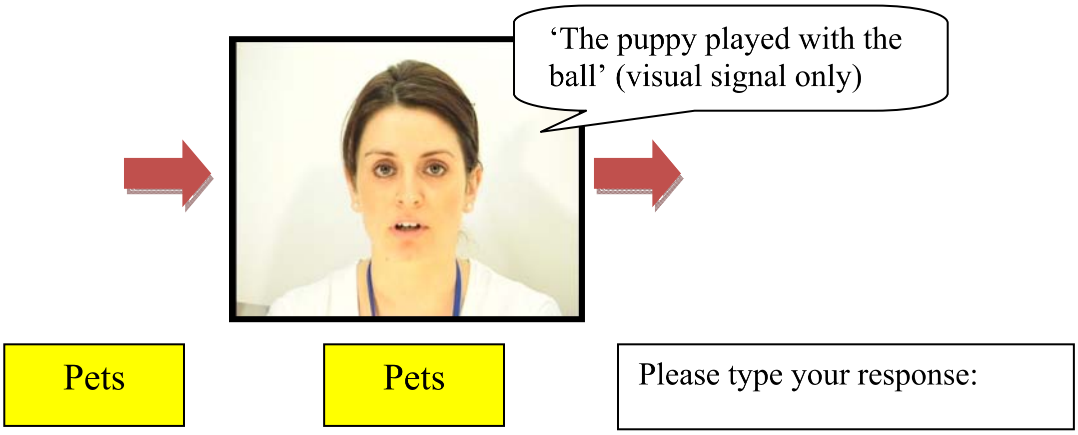


Figure 5.3: Diagram of experimental procedure

5.2.2 Results

The results were analysed initially using a paired t-test which showed that there was a significant decrement in speechreading accuracy associated with the Glaswegian accent (mean keywords correct: 42.27 = 12% correct) compared with the Nottingham accent (mean keywords correct: 64.93 = 18% correct), despite the use of a contextual cue: $t\ (14) = -5.569$, $p < 0.05$. However performance did appear to be generally improved, with scores ranging from 12 to 18% correct across accent types in comparison to 3 to 7% correct from the same set of sentences in Experiment 5, where no contextual cue was present.

To examine whether the contextual cue provided a significant benefit, the present data were combined with data from Experiment 5 (Nottingham participants' data only) to allow comparison of speechreading performance on the same sentence sets, using the same talkers, but with or without a contextual cue. Analysis using a two-factor, mixed factorial (2 context x 2 accent type) ANOVA indicated that, as before, there was a significant effect of accent upon the results: $F\ (1, 37) = 62.788$, $p < 0.05$. However, there was also a significant effect of context upon performance: F (1, 37) = 32.974, $p < .05$. This suggests that although context did not reduce the effect of regional accent upon performance by reducing the difference in performance across the two accent types, there was a significant increase in scores for both accents. Thus, contextual cues appear to improve the visual intelligibility of the Glaswegian talkers, but do not raise speechreading performance to the level achieved when participants are presented with Nottingham talkers. Finally, there was no interaction between the two factors: F (1, 37) = .810, $p > 0.05$. This indicates that the poor performance associated with the Glaswegian accent type was of the same magnitude in both the context (Experiment 10) and no-context (Experiment 5) conditions, as illustrated by Figure 5.4.

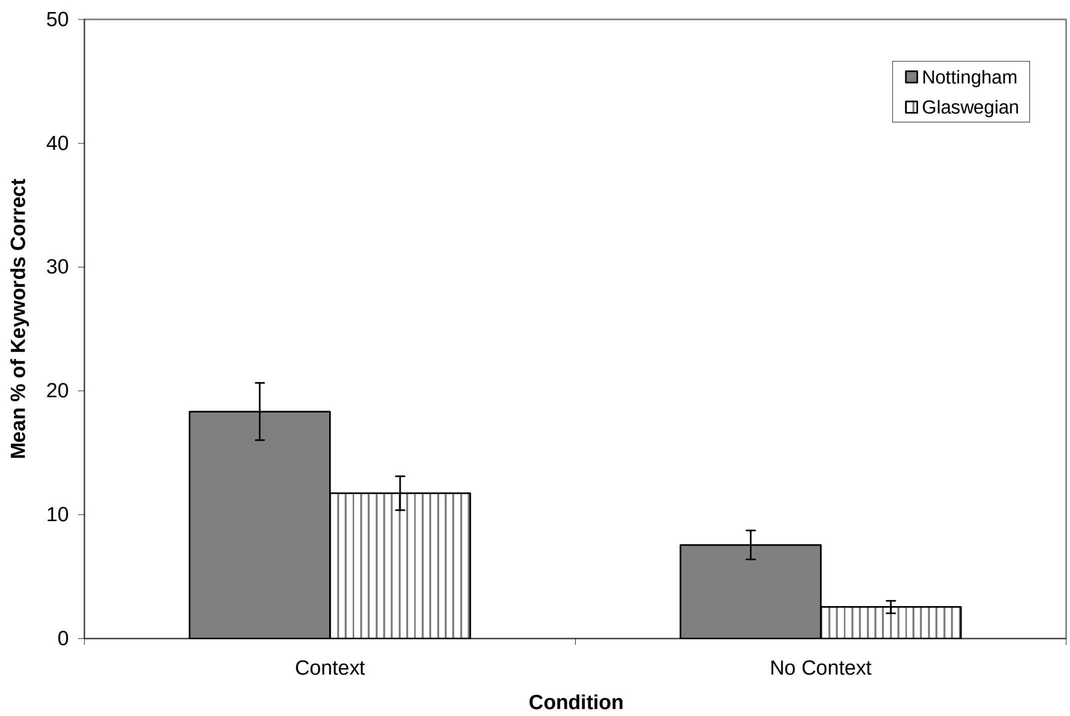


Figure 5.4: Mean Percent Keywords correct for context and no-context conditions, shown with standard error bars.

Figure 5.4 illustrates the benefit in performance provided by the use of a contextual cue, indeed a post hoc independent t-test comparison of the performance level recorded for the Glaswegian talkers when a contextual cue was present and the Nottingham talkers when no cue was present, showed a significant difference between the two scores, $t(37) = -2.280$, $p < 0.05$. This indicates that use of a contextual cue can improve the intelligibility of the Glaswegian accent to a level which is higher than that recorded for the familiar Nottingham accent when no cue is present.

The next step in the analysis was to determine if the recorded increase in performance due to context was equal across all talkers, to do this the results were plotted on a talker-by-talker basis, as illustrated by Figure 5.5.

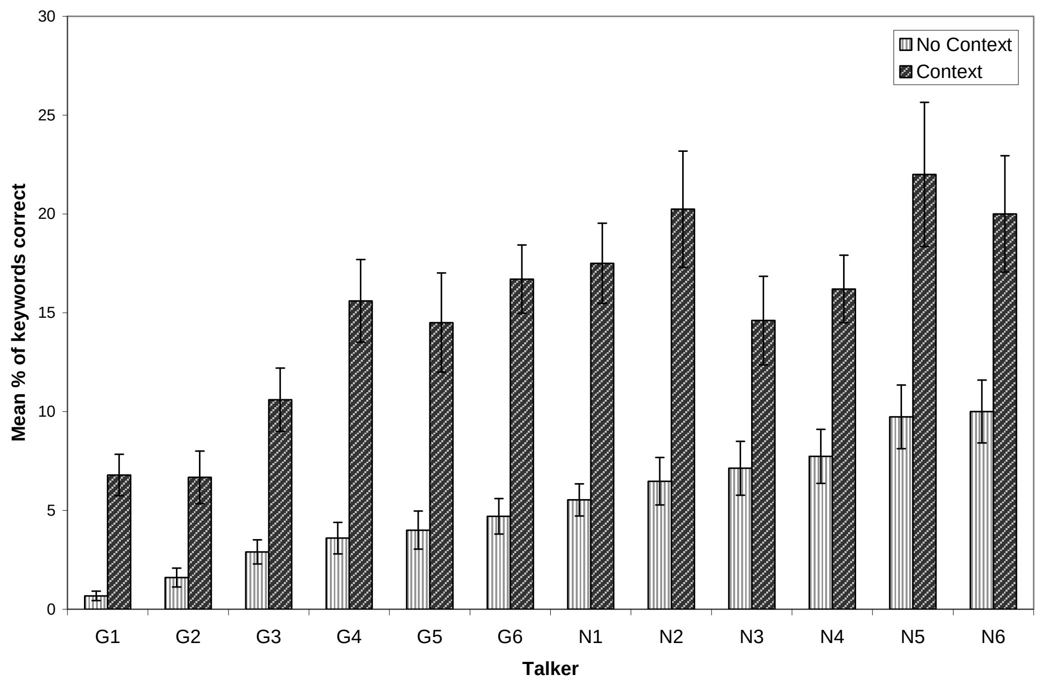


Figure 5.5: Mean number of keywords correctly identified by talker and context condition, shown with standard error bars.

Analysis of these results using a mixed factorial (12 talkers x 2 context / no context) ANOVA found that there was a significant effect of talker: $F(11, 407) = 28.149$, $p < 0.01$. There was also a significant effect of context: $F(1, 37) = 133.927$, $p < 0.01$. This indicates that contextual cues did significantly improve performance across the talkers. Finally, there was also a significant interaction between talkers and context: $F(11, 407) = 3.640$, $p < 0.01$. This indicates that the benefit given by the contextual cues varied across the talkers. Post-hoc analysis of the data from the Experiment 10 using Bonferroni threshold correction for multiple comparisons indicated no significant difference between Glaswegian talkers (G4, G5 and G6) and the Nottingham talkers (N3 and N4). This suggests that context has improved some of the Glaswegian talkers above others, raising their speechreadability to a level on a par with some of the Nottingham talkers (for the full analysis see Appendix 6). This indicates that context yields a different strength of effect depending on factors associated with the talker's inherent level of speechreadability, namely that certain facets of speech production may be enhanced through the generation of expectations where others may not.

5.2.3 Discussion

The results indicate that a) context improves speechreading performance significantly and b) although the scores for Glaswegian talkers were more accurate than previously, there still remains a significant performance cost for an unfamiliar accent. These results illustrate the potential usefulness of context in a conversational setting, providing the observer with linguistic constraints that aid them in their perception of speech. Our findings also indicate that lack of familiarity with, and certain characteristics of, an accent type can still impair speechreading accuracy, illustrating its importance as a factor in speech perception.

The observed general improvement in performance demonstrates the successful use of topical constraint by the participants as a speechreading aid. This is in agreement with past research on the subject, already detailed in this chapter (Lansing & Helgeson, 1995; Lidestam et al, 2001). It has been suggested that measures of speech perception utilising context may be a more accurate method of testing, being closer to the natural flow of conversation than sentences or words produced in isolation (Flynn & Dowell, 1999). Certainly, conversation within an everyday setting should allow the average speechreader to utilise a variety of cues, including linguistic redundancy. These cues allow the speechreader to 'fill in the gaps' left by the less visible aspects of articulation (Boothroyd, 1988) hence the general increase in speechreading performance. However, as reported here and in earlier studies of contextual effects on visual speech (Lansing & Helgeson, 1995) the visual intelligibility of a talker also has an influence upon the results. There are two effects that should be discussed in further detail; accent effects and talker variability.

First, although there is no doubt that the use of context led to an improvement in performance for the Glaswegian talkers, the continued difference in performance between the two sets of speakers indicates that accent effects persist even when accuracy is improved. This was surprising when compared to past research which indicated that contextual cues were more advantageous when task difficulty increased (Lidestam et al, 2001). The effect of context upon the Glaswegian speakers was therefore expected to be proportionally larger than the improvement for the Nottingham talkers. Since this was not the case, it suggests that the articulatory gestures associated with the Glaswegian accent are resistant, to a certain extent, to the aid given through contextual cues. Erber (1992) posits that observers develop a strong set of expectations based on contextual cues and that when these expectations are not fulfilled the observer is unable to prescribe meaning to the visual speech that follows. Thus, the contextual cues may have been effective for the visual speech produced in the Nottingham accent because the viewers were familiar with that accent and had their expectations of the appearance of visual speech fulfilled by the familiar visual cues. In comparison, the visible articulations associated with the Glaswegian accent were comparatively unfamiliar, increasing the likelihood of misperception.

Auditory research utilising context to reduce accent effects (Labov, 1989; Labov & Ash, 1997), reported similar results to those found in Experiment 10, with accent effects proving fairly resistant to contextual cues. Labov (1989) suggests that the unfamiliar phonetic form can essentially ‘block’ accurate perception of the message. That is, even when the context given clearly indicates the correct interpretation of the stimulus, the listener is unable to ignore the unfamiliar pronunciation, leading to misinterpretation of the stimulus. Thus, it is possible that the unfamiliar articulation patterns used by some of the Glaswegian speakers ‘blocked’ correct interpretation of the stimulus producing a similar effect. This result may also relate to the discussed N400 / P600 effect (Connolly & Phillips, 1994; Calabresi, 2007), the N400 effect results from the perception of a semantically incongruent word at the end of a sentence, resulting in increased processing to interpret its meaning. Within visual speech the response has been recorded as a P600, suggesting a stage of reanalysis of the stimulus (Calabresi, 2007). It is possible that the Glaswegian accent induced misperception of the visually presented sentences, which may have resulted in the observer classifying the viewed words as incongruent to the contextual cue provided. This could have resulted in the observer reanalysing the stimulus, reducing the potential advantage provided by the contextual cue.

Variation was also present in contextual effects across the talkers of both accent groups. This suggests that there are aspects of talker intelligibility other than accent that have an influence on the level of improvement produced by context. Prior research indicates that context can improve the segmentation of the auditory speech signal (Cole, Jakimik & Cooper, 1980) and compensate for some aspects of low visibility that relate to (high) word frequency and labial articulations (Lansing & Helgeson, 1995). It is probable that the variation in performance improvement across talkers is linked to the relative word visibility and segmentation of speech produced by each individual talker's visible articulation. This indicates that talker factors which might negatively impact visual speech production, such as lip shape, rhythm and viseme production are unaffected by contextual cues, producing the variation in context related improvements shown within this study.

To conclude, context increases speechreading performance through contextual constraints, but does not mitigate regional accent effects entirely. In order to reduce those accent effects further, it may be necessary to increase exposure to a particular accent type in order to allow the observer to familiarise themselves with the associated articulatory movements.

5.3 Experiment 9: The impact of contextual constraints and increased exposure upon regional accent effects on speechreading performance

Accent effects seem resistant to increased exposure or contextual constraints, although contextual cues do result in a general improvement in performance. The aim of Experiment 9 was to examine increased exposure to the Glaswegian accent, with the provision of context to improve general speechreading performance. As discussed previously, it is possible that the unfamiliar articulations of the Glaswegian talkers reduced the advantage provided by contextual cues, producing the recorded detrimental effect of the Glaswegian accent upon speechreading accuracy. By increasing the exposure level of participants to the articulatory motions of four Glaswegian talkers, we hope to induce further improvements in performance by familiarising them with the accent type. Past research has shown an improvement in speechreading performance through talker familiarity, once floor effects were removed using a low level of speech sounds (~20 dB at 2m) (Lander & Davies, 2008). We hope to expand that effect to include accent familiarity by removing floor effects through the utilisation of context

The experimental hypotheses were as follows:

- The combination of increased familiarity with the talker's facial movements, together with contextual constraints, should improve speechreading performance for the Glaswegian accent.
- This improvement should bring speechreading performance for that regional accent up to the same level as the observed performance for the Nottingham accent.

5.3.1 Method

Participants

Eight participants were recruited for the study, all were native English speakers from Nottingham and reported a good level of hearing and normal (or corrected to normal) vision.

Stimuli

In total, 340 BKB sentences were recorded using eight talkers, four of whom had a Glaswegian accent, the remaining four having a Nottingham accent. Of those sentences 160 were recorded for use as pre- and post-tests of speechreading ability (80 sentences for each test), the remaining 180 were recorded for use as the training session. As before, each sentence set was balanced across talkers by matching sentences for visual content. To further counterbalance any remaining effects due to item difficulty, the order in which the test sessions were presented to each participant was randomised, hence participant one might view test sessions 1 then 2 in that order, then participant two would view the test sessions in the order of 2 then 1 and so on. The training sessions were the same for all.

Each talker's face was fully illuminated and recorded against a light background with only the face and neck visible for each sentence recording.

Procedure

The experiment was made up of three parts: pre-test, training and post-test. Each test session consisted of 80 BKB sentences presented visually. These comprised eight sets of ten sentences, each set spoken by a different talker (four

Nottingham and four Glaswegian talkers). The training session contained 180 sentences, split into four sets of 45 sentences, each set produced by one of the four Glaswegian talkers (same as test session). Presentation of the talkers was randomised across all test and training sessions.

Each participant was seated at a table directly in front of the view screen and instructed to watch each clip carefully. They were instructed that a cue word would appear on screen, followed by a silent video clip of a talker producing a short sentence. Their task was to speechread the sentence presented and then type in any words they had seen on the computer. Performance was measured by comparing the number of keywords correctly identified within each sentence set.

5.3.2 Results

As before, each participant's score represents the number of keywords they correctly identified, marked using the loose scoring criterion. On this occasion a participants score was calculated from a possible 240 keywords per test session. Initial viewing of the results indicated a small drop in performance for the Glaswegian talkers after training (pre-test = 9%, post-test = 8% keywords correct), although performance was higher than that reported in Experiment 7 (pre-test = 1 – 2%, post-test = 1 – 3% keywords correct). Performance levels for the Nottingham talkers in Experiment 9 remained higher (pre-test = 12%, post-test = 10% keywords correct) than the results for the Glaswegian talkers. This indicates that contextual cues improved general levels of performance but did not increase the effectiveness of exposure in reducing accent effects.

This initial observation was confirmed by analysis using a repeated measures (2 accent type x 2 test session) ANOVA, the mean scores for the two test sessions were shown to differ significantly: $F(1, 7) = 11.413$, $p < 0.05$. The difference between mean scores for the two accent types were also found to be significant: $F(1, 7) = 9.779$, $p < 0.05$. Finally, there was no significant interaction between the two factors: $F(1,7) = .144$, $p > 0.05$. Figure 5.6 illustrates these results.

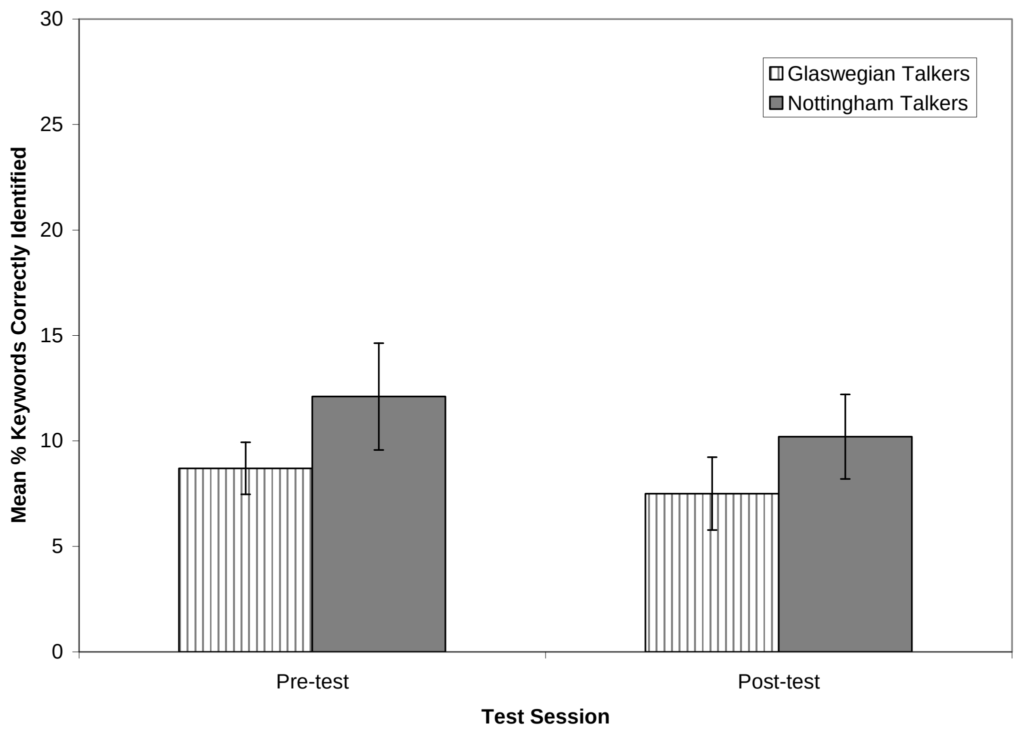


Figure 5.6: Mean percent keywords correct for Nottingham and Glaswegian talkers across two test sessions, shown with standard error bars.

5.3.3 Discussion

Previous visual research has indicated that a period of familiarisation with a talker significantly improved speechreading performance for that talker (Lander & Davies, 2008). The authors suggest that a relatively short period of familiarisation (2 minutes) is sufficient for an observer to 'tune in' to the talkers speaking style and consequently improve that talker's visual speech intelligibility.

A similar result was expected here, but generalised across an accent type. Thus, rather than familiarising our participants with a single talker, we attempted to familiarise them with four talkers, all with the same Glaswegian accent. Contrary to the study by Lander and Davies (2008) the results from the present study indicate that increased exposure to several talkers does not improve speechreading performance, despite the removal of floor effects using contextual constraints.

The lack of a significant improvement in speechreading performance for the Glaswegian talkers may be due to talker variability effects. The study by Yakel and colleagues (2000), and the results shown by Experiment 6 in this thesis, indicate that speechreading performance is reduced through the use of multiple talker lists (see Chapter 4, Experiment 6 for a full discussion). Thus, in the present study it is possible that the processing requirements for adapting to the speech of multiple talkers reduced any positive effects associated with increased exposure and contingent familiarity effects. If the experiment was repeated using a single talker, thereby requiring the visual system to adapt to a single speaking style and accent type, it is possible that a more positive result would be produced. However, in comparison to the results reported here, research within the auditory domain has reported accent familiarity effects using multiple talkers (Floccia et al, 2006). Thus, the results of the present experiment suggest a relatively inflexible visual system, unable to swiftly adapt to the visual variance produced by an unfamiliar, or difficult, accent type when faced by multiple talkers. The auditory system is comparatively more flexible, showing significant improvements in understanding accented speech when multiple (Floccia et al, 2006) or single (Clarke & Garrett, 2004) talkers are used.

These results also reinforce the suggestion that the Glaswegian accent may contain elements which are inherently difficult to speechread, regardless of context or continued exposure. It is possible that the Glaswegian accent reduces the visibility of various articulations, thereby restricting talker intelligibility and reducing any possible improvement through exposure and contextual cues. The specific nature of this effect is outside the framework of this thesis and would therefore benefit from further research in this area.

Finally, the results show a decrease in performance between the pre- and post-test measurements of keyword accuracy. The reason for this decrease is unclear, but is most likely due to attentional deficits associated with prolonged testing periods (Experiment 9 was approximately 45 minutes long). Also, although significant, the fall in performance is relatively small (approximately 2%) and must be judged on the relatively small number of participants recruited for this study (8 participants). Thus, it is possible that this result is due to sampling error and would not be present within a larger sample of the population.

To summarise, the results again confirm that the visual speech produced by Glaswegian talkers is less intelligible to Nottingham observers than their 'home' accent. This reduced intelligibility remains despite efforts to reduce the detrimental effect through familiarising the observers with the talkers using exposure or through the utilisation of contextual cues. This suggests that the visual speech system is less able to cope with accent variation than its auditory counterpart. However, the improvement in performance produced by contextual cues indicates that under normal everyday conditions a speechreader may be more accurate at deciphering the visual signal through their use of conversational cues than our participants in a controlled laboratory setting. This led to the consideration of other strategies potentially used by speechreaders to improve visual speech intelligibility. One such tool is repetition (Palmer, 1988); asking a conversational partner to repeat themselves when an observer has failed to comprehend their message. The repetition 'repair mechanism' (clarification of a misperceived message) is the favoured mechanism used by hearing-impaired adults (Marzolf, Stewart, Nerbonne & Lehman, 1998) to improve speech intelligibility, it is for this reason that we included it within Experiment 10.

5.4 Experiment 10: The effect of repetition priming on the impact of an unfamiliar regional accent upon speechreading performance

Repetition in an everyday setting involves asking a fellow communicator to repeat an unclear phrase or sentence, and is a common tool used by hearing-impaired individuals to facilitate comprehension (Palmer, 1988). The repetition offers a second chance at speechreading, while the first view provides priming for the second production of the message. Essentially two theories account for repetition priming, in both auditory and visual speech. The first is abstractionist and states that prior contact with a word (or sentence) facilitates later processing through activation of the item's lexical representation (Bodner & Masson, 1997). Thus, priming results in the ongoing activation of lexical entries, which then speed later classification or comprehension of the stimulus when it is repeated a second time (Dennis & Schmidt, 2003). The second is episodic and states that priming is the result of the primary encounter with a stimulus being encoded into memory. This information is then retrieved when the stimulus is encountered a second time, speeding processing time (Dennis & Schmidt, 2003). Despite the differences between the two theories, both posit a positive effect of repetition of a stimulus.

This facilitatory effect has been previously examined within several disciplines including, masked repetition priming of written words (Bodner & Masson, 1997) where lexical decision-making reaction times were reduced through priming effects. Cross-modal priming effects for target words in naming and lexical decision-making tasks have also been examined (Kim, Davis & Krins, 2004), with the finding that visually presented words produced a reliable priming effect for auditory and written decision-making tasks. The authors suggest this indicates that visual speech can give rise to the same lexical activation as provided by auditory speech, illustrating the amodal nature of priming effects. Finally, the threshold of perception for auditory speech in noise was found to be improved by a shift of 2 dB (Miller, Heise & Lichten, 1951) when the test item (word) was repeated. This indicates that the repetition of an item improves speech intelligibility, thus reducing the masking effect of noise upon speech.

There have also been several studies examining repetition priming as a repair strategy within the domain of visual speech. For example, Squires and Dancer (1986) investigated the effect of repetition upon visual speech and auditory speech-in-noise. Participants were asked to transcribe 100 words presented as either acoustic stimuli in noise or video recordings of visual speech, in both cases the entire list was repeated a second time immediately after the first presentation. The results showed that performance increased significantly for the second presentation of both types of stimulus, by 11% in the auditory condition (from 48 to 59% correct) and by 3% (from 25 to 28% correct) in the visual condition. The authors argue that these increases in performance are equivalent to those provided by other training mechanisms (such as feedback) and as such should be considered in any speech training programme (Squires & Dancer, 1986).

In a similar study repetition of several different types of visual stimulus (words, phrases and sentences) were found to significantly improve speechreading accuracy (Ijsseldijk, 1992). Three types of repetition were considered for each type of stimulus – entire face, profile and lips only. In each case the level of improvement was approximately 4% (average performance increase = 29% to 33% correct). The authors report that the method of repetition; whether it involves the full face or part of the face, is irrelevant as in each case an improvement in performance was produced (Ijsseldijik, 1992).

Finally, a further study compared the effect of five repair strategies; repetition of entire stimulus, simplification (repetition of the sentence in a simplified form), rephrasing, repeating a keyword and producing the sentence as two separate sentences, each containing further information, upon the intelligibility of visual sentences (Tye-Murray, Purdy, Woodworth & Tyler, 1990). The results, as shown in Figure 5.7, indicate that the improvement associated with each type of repair mechanism was equivalent.

This indicates that the effectiveness of repetition as a strategy to improve visual speech intelligibility is at least as effective as the other mechanisms utilised by hearing-impaired or deaf speechreaders. Interestingly, Tye-Murray et al. (1990) did not find any effect of talker (six talkers were used to produce the sentences, accent type unknown) on the benefits provided by the repair strategies. In terms of the present study, this indicates that repetition may improve the intelligibility of talkers irrespective of accent type.

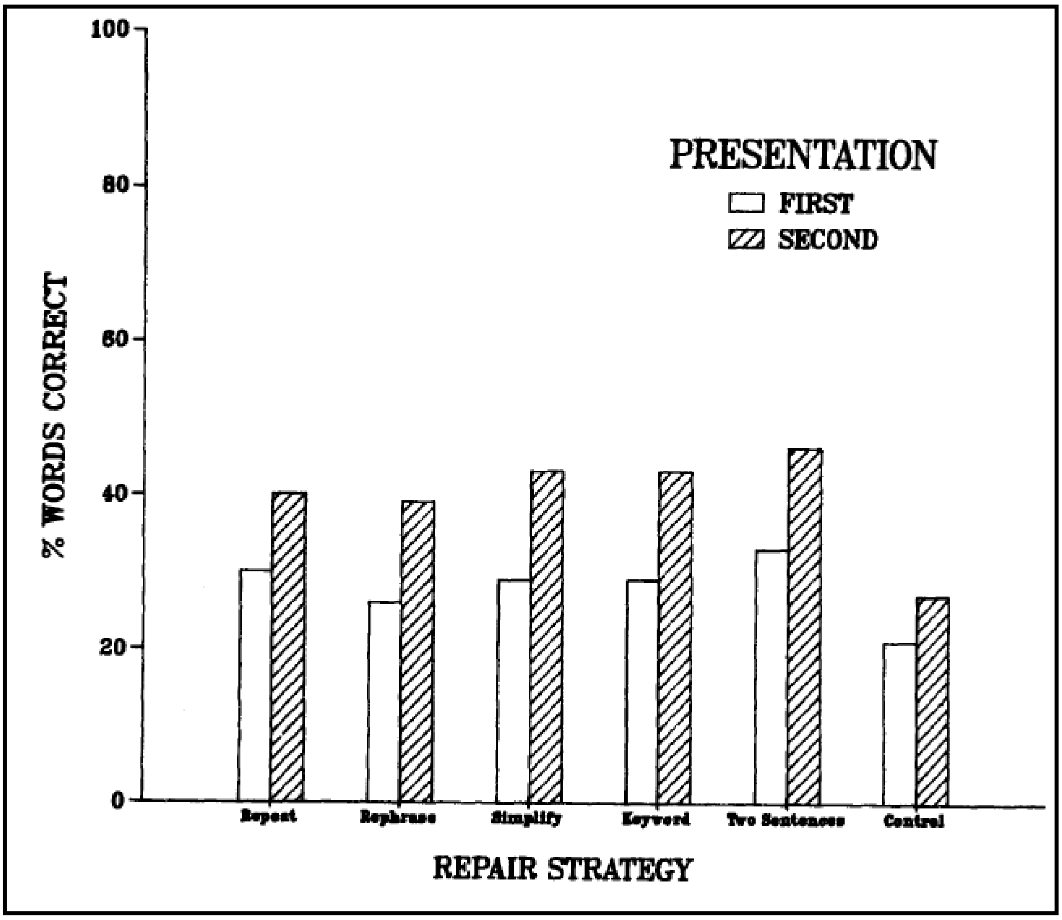


Figure 5.7: The averaged first and second presentation scores (Tye-Murray, Purdy, Woodworth & Tyler, 1990).

The research reported above indicates the effectiveness of repetition as a method by which to improve visual speech perception. The purpose of the present study is to examine the potential increase in performance provided by the repetition of a visual sentence, with a particular focus upon any improvement in the intelligibility of talkers with a Glaswegian accent. Essentially, we propose that a combination of contextual cues (already shown to have a facilitatory effect upon visual speech intelligibility in Experiment 8), and repetition (previously shown to improve intelligibility irrespective of talker variability) may together reduce accent effects. The experimental hypotheses are thus:

- Repetition of a visual sentence should improve general visual speech intelligibility for both talkers with a Nottingham accent and those with a Glaswegian accent.
- The combination of contextual cue and repetition should elicit a larger improvement in general performance than the use of contextual cues alone.
- The use of contextual cues and repetition together should reduce accent effects through their improvement of the intelligibility of talkers with a Glaswegian accent.

5.4.1 Method

Participants

Fifteen participants were recruited for the study, all were native English speakers and reported good hearing and normal (or corrected to normal) vision. All of the participants had been born in the East Midlands area and had lived in Nottingham for more than two years.

Stimuli

A total of 240 sentences from the BKB (Bench, Kowal & Bamford, 1979) were used. The sentences were split into 12 lists comprising ten sentences each, with every sentence being shown twice in immediate succession producing a total set size of 20 sentences, each set spoken by a different talker.

The talkers used for the recordings were grouped as follows: six talkers (three male, three female) with Nottingham accents, while the remaining six (four male and two females) had Glaswegian accents. All of the stimuli are identical to those used in the experiments described prior to this. As before, each sentence was preceded by 1s of the talker's static face, and followed by a further 1s of static recording. Each clip contained only visual information.

<u>Procedure</u>

Participants were seated at a table directly in front of the computer screen, and instructed that they would first be shown a 'cue word', described as providing a clue as to the nature of the sentence. After the cue word had appeared a video clip would follow shortly after. They were further instructed that the talker would produce one sentence per video clip, which they were asked to watch carefully. Their task was to identify what the speaker had said and type their response into a keyboard. They would then view the same sentence a second time and respond again. They were not required to understand the entire sentence; any word that was typed in was recorded. All fifteen subjects viewed the clips presented in a randomised order.

5.4.2 Results

Again, each participant's score represents their accuracy at the test – that is the number of keywords correctly identified within a sentence set (maximum number correct = 360 for each set) scored using the loose scoring criterion. Initial analysis of the results indicated that repetition of the stimulus increased performance levels, with scores for the Glaswegian talkers rising from 10% (mean 19 keywords) correct to 14% (mean 26 keywords), a 4% increase in correctly identified keywords for the repeated stimulus and from 16% (mean 29 keywords) to 20% (mean 37 keywords) a 4% increase, for the Nottingham talkers. Interestingly this meant that final scores for the second repetition were higher than those recorded when contextual cues alone were utilised in Experiment 8 (12% correct for Glaswegian talkers, 18% correct for Nottingham talkers), indicating that repetition did further improve performance when combined with contextual constraints. However, a comparison of general performance using an independent t-test indicated that this increase in performance between the two experiments (Experiments 8 and 10) was not significant: $t(58) = -.234$, $p > 0.05$.

The observed increase in performance across the two presentations of the stimulus in Experiment 10 was confirmed using a repeated measures (2 accent x 2 repetition = 1st viewing versus 2nd viewing) ANOVA. The results showed that there was a significant main effect of accent: $F(1, 14) = 15.888$; $p < 0.05$. There was a significant main effect of repetition: $F(1, 14) = 25.639$; $p < 0.05$. Finally, there was no significant interaction between the two factors: $F(1, 14) = 0.692$; $p > 0.05$. These results are illustrated in Figure 5.8.

The results indicate that repetition of the stimulus led to a significant increase in performance for both sets of talkers, but that the cost of the Glaswegian accent upon speechreading accuracy still remained when the second presentations of the stimuli were compared.

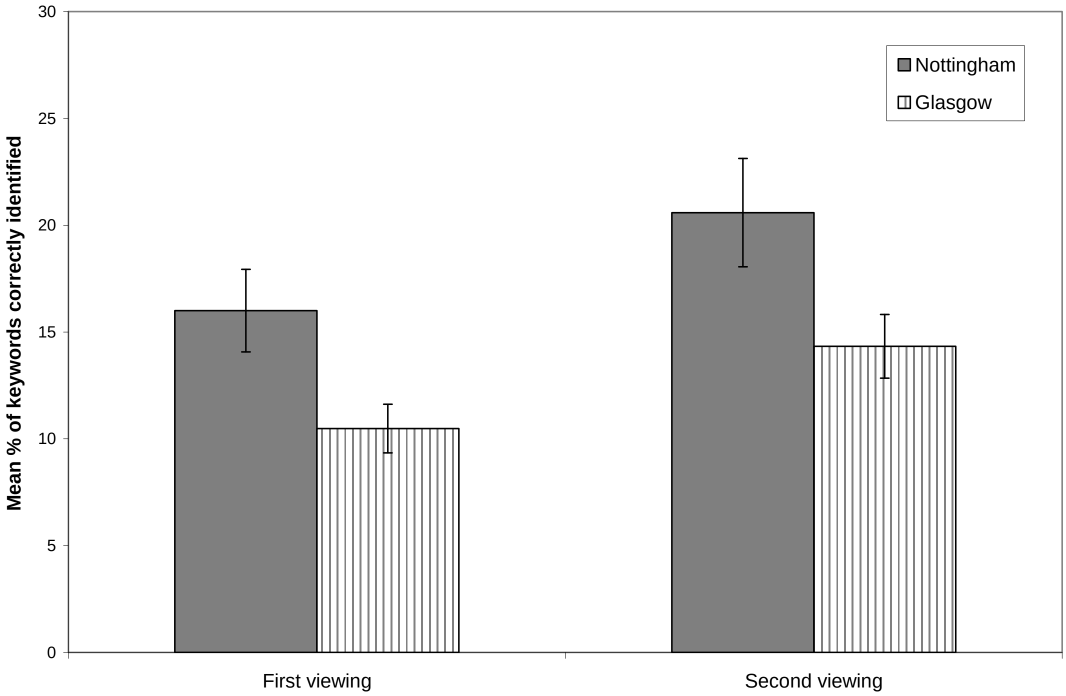


Figure 5.8: Mean percent keywords correct for first and second (repetition) viewing of Nottingham and Glaswegian talkers, shown with standard error bars.

However, of interest was a comparison of speechreading performance between the first presentation of the Nottingham stimuli and the second presentation of the Glaswegian stimuli. This was to determine if, in an everyday setting, asking a talker with an unfamiliar or difficult accent to repeat themselves would result in intelligibility for that talker increasing to a level comparable with the first presentation of a talker with a familiar or easier accent type. Analysis using a paired t-test revealed that there was no significant difference in performance between the second presentation of the Glaswegian talkers and the first presentation of the Nottingham talkers: $t(14)$: -1.237, $p > 0.05$. This indicates that the repetition of the Glaswegian stimuli increased performance to a level equivalent with the first presentation of the Nottingham talkers. This represents the first indication that repair mechanisms such as repetition and the use of contextual constraints may be able to compensate, to some degree, for accent effects.

5.4.3 Discussion

The results confirm previous research in this area (Tye-Murray et al, 1990; Squires & Dancer, 1986) by illustrating the effectiveness of repetition as a repair mechanism capable of improving visual speech intelligibility. The improvement was larger than that found in Experiment 8 for contextual cues alone, but this difference was not significant. This lack of significance should be viewed with caution due to potential sampling differences for the two groups of subjects; it is possible that with a larger sample of the population this result would become significant. Furthermore, in the Tye-Murray et al. (1990) study, they repeated the stimulus three times in order to procure the equivalent increase in performance to the other repair strategies used (paraphrasing, simplification etc.). It is possible that a third repetition in this case would have improved performance significantly more than contextual cues used in isolation. However, based on the present results, it does appear that contextual constraints are more effective at facilitating speechreading performance than the repetition repair strategy. This suggests that the provision of increased information as to the content of a message is more helpful than simply viewing the stimulus a second time.

The detrimental cost associated with the Glaswegian accent persisted even when both contextual constraints and repetition were utilised, illustrating the importance of accent in the relative speechreadability of a talker. Certainly, such a result indicates that there are aspects of the Glaswegian accent that do not benefit continued exposure, contextual constraints and repetition. This result disagrees with previous research (Tye-Murray et al, 1990) which found benefits associated with repetition to be independent of talker characteristics. However, no mention was made of accent type in their descriptions of the talkers used, so it is assumed that the talkers had similar accent types. A comparison with the present results indicates that accent as a factor of talker variability has a greater influence on the effectiveness of repair strategies than other associated talker characteristics, such as lip shape (Berger, 1972). This relative importance of accent type could be due to either the paucity of the visual signal associated with visible characteristics of the Glaswegian accent, or the processing strategy of the visual system when dealing with accent variation. Further research is required in this area in order to determine which possibility has the greatest influence.

As stated, a direct reduction in accent effects was not apparent when the conditions of stimulus presentation between the two accent types were equal. However, the comparison of performance between the first presentation of the Nottingham stimuli and the second presentation of the Glaswegian stimuli gave an indication of the potential usefulness of repetition in an everyday environment. The increase in performance associated with the second viewing of the Glaswegian stimuli improved the intelligibility of the talkers to a level equivalent with the first viewing of the Nottingham talkers. This is our first indication that various repair strategies might be able to compensate for some of the effects associated with accent variation. Certainly it's possible that, within a conversational setting, speechreaders may utilise repair mechanisms such as repetition to improve their understanding of unfamiliar or accented speakers. Such a strategy should reduce everyday accent effects to some degree.

Despite the usefulness of repetition in potentially decreasing the impact of accent upon visual speech intelligibility, the general persistence of accent effects throughout this chapter indicates that it is a factor which should be considered when developing a test of speechreading performance. Previous research (Demorest & Bernstein, 1992) has already indicated that talker variability is an important factor for consideration when evaluating speechreading performance. The results of this chapter indicate that accent type is an important aspect of talker variability and as such could affect the comparison of speechreading scores across tests utilising talkers with different accents. For that reason accent type should be considered when developing a test of speechreading ability.

5.5 General Summary

To summarise, the research shown within this chapter indicates that the visible characteristics of an accent type can influence the intelligibility of a talker. This effect is not reduced through increased exposure to talkers with that accent type or contextual constraints. The use of contextual cues has been shown to improve the speechreadability of talkers with the Glaswegian accent, but that improvement is not greater than the comparable increase in performance for talkers with a Nottingham accent. Finally, a combination of repetition and contextual cues appear to be the most effective when compensating for accent effects in visual speech.

Chapter 6: General Discussion and Conclusions

The research reported in this thesis investigated the effect of accent type upon visual speech perception using behavioural methods. This was motivated by a fundamental lack of research on these effects within the visual speech domain, whilst concurrently a large body of research existed on the topic within the auditory research domain.

6.1 Research aims

The influence of general talker variability on speechreading accuracy is a well-developed area of research within the visual domain. However, although various factors have been considered (lip shape, speech rate etc. see Lesner, 1988; Massaro, Cohen & Gesi, 1993) there is still relatively little known about the specific parameters of talker speechreadability. Furthermore, although research exists that details the potential impact of observer-based characteristics on speechreading performance (Conrey & Gold, 2006), there has been no examination of a possible interaction between the observer and talker factors. The influence of accent on the visual signal could incorporate both talker parameters (the influence of accent characteristics upon talker speechreadability) and observer factors (the impact of observer generated expectations on the processing of visual speech) and thus represents an important aspect of visual speech and one which has both theoretical and practical ramifications. First, theories of speech perception split into two general categories: 'encoding' and 'normalisation'. The application of these theories to auditory research describe the ability of a listener to compensate for talker variation in the speech signal, either through screening out talker characteristics as 'noise' or through encoding talker information as a valid aspect of speech production. Implicit in these theories is the suggestion that a similar process will occur within visual speech perception. Moreover, the study of accent within the auditory domain suggests that a talker's accent type is an important aspect of variation in speech production and that accent familiarity constitutes an integral part of speech comprehension. Thus, the study of accent effects in the visual modality constitutes an examination of the applicability of auditory theories of speech perception to visual speech and a comparison of talker-specific characteristics across the modalities. Second, improved understanding of the factors that influence talker speechreadability could both aid in the production of more ecologically valid tests of speechreading performance and potentially help deaf individuals to maximise the effectiveness of speechreading training measures. The thesis therefore had three key aims. First, to determine whether accent represents a salient aspect of the visual speech signal. This was measured through the utilisation of visual cues by an observer in order to determine accent type from the visual signal. Second, to investigate the effect of accent on speechreading performance. This encompassed analysis of the influence of accent type upon talker speechreadability, the effect of accent familiarity on speechreading performance and an investigation into accent variation. Third, the reduction of accent effects by manipulating the stimulus. This included the utilisation of continued exposure to an accent type, the addition of contextual cues to a stimulus and the repetition of a stimulus.

6.2 Summary of main findings

Figure 6.1 illustrates the main findings of this thesis by updating the theoretical framework first introduced in Chapter 1, Figure 1.1. The changes to the model are illustrated by blue lines and boxes.

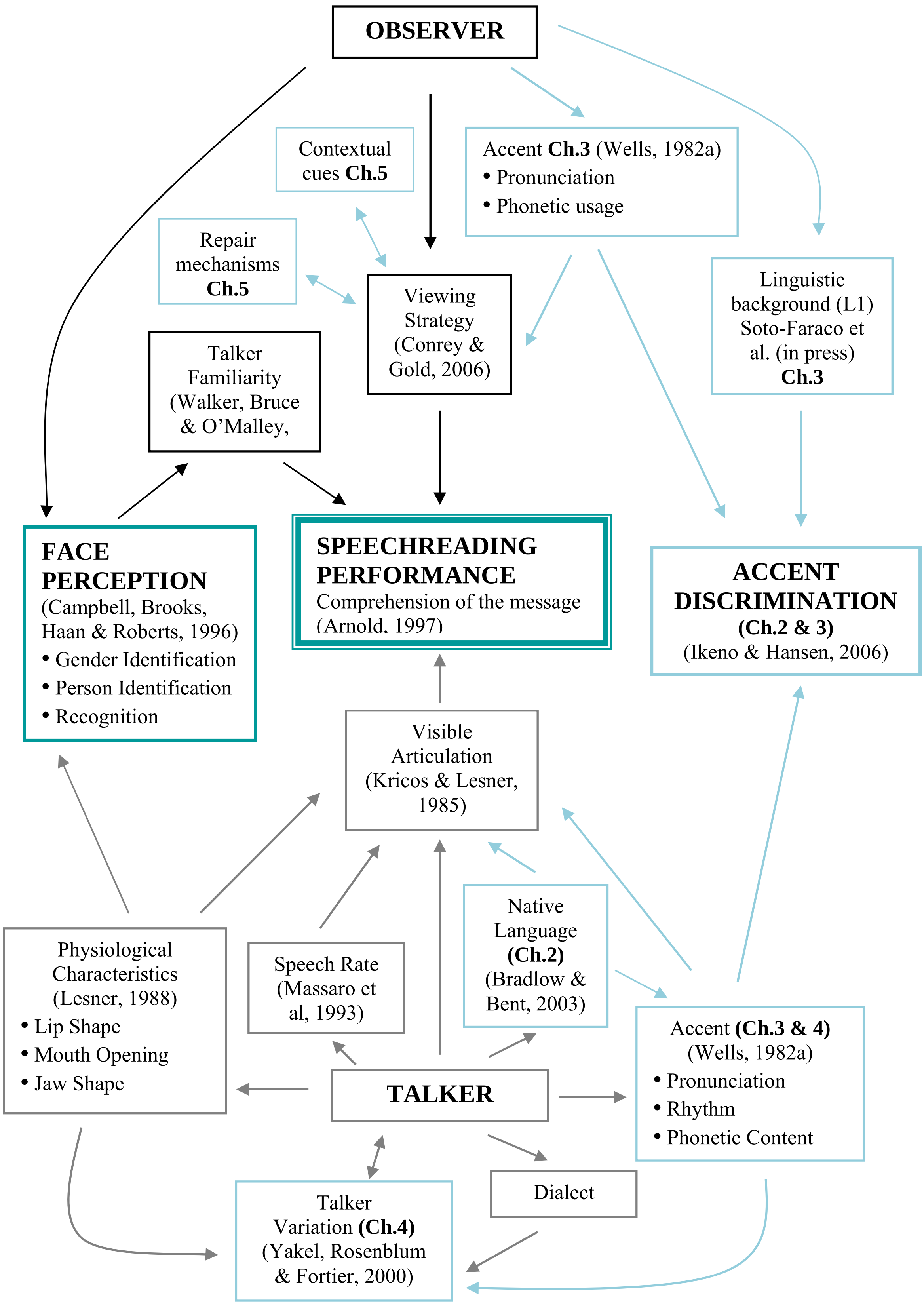


Figure 6.1: Detailed framework showing how speechreading performance (centre) is determined by observer (top, black lines) and talker (bottom, grey lines) characteristics, blue lines represent results found within this thesis

6.2.1 Chapter 2: Intra-talker foreign accent variation

Two aspects of intra-talker foreign accent manipulation were examined. First, the ability of observers to utilise visual cues to discriminate accent and language type (French/English) was analysed. Second, the influence of foreign (French) accent production upon talker speechreadability was investigated. The results showed that observers were able to discriminate between the two languages and accents by utilising visual cues at a level significantly above chance. However, incongruent stimuli in which the language and accent did not match had an adverse effect on discrimination performance. Furthermore, the manipulation of the talker's accent was found to have an adverse effect upon his speechreadability, with a significant decrement in observer speechreading performance associated with his French-accented utterances. The results from this study constituted the first indication that accent type was both a salient aspect of the visual signal and a factor which could impact talker speechreadability.

6.2.2 Chapter 3: Inter-talker regional accent variation

The first aspect of inter-talker regional accent variation to be examined was the discrimination of Nottingham and Glaswegian accents using either the visual or auditory modality. There were two main findings. First, accent discrimination was possible in both modalities, but performance was significantly more accurate in the auditory modality. Second, linguistic experience was a constraint on performance levels, with non-native English speakers significantly less accurate at the task. Individual variation in both tasks was higher within the non-native compared to the native English speaking participants (see Chapter 3, Experiment 4B results). This was attributed to the use of compensatory strategies by the non-native speakers since their judgments were based on restricted knowledge of an L2. The native participants were much more practiced at recognising accent cues within the English language auditory signal and consequently exhibited both higher performance levels and reduced individual variation as scores approached ceiling level. However, native participant scores for the visual modality were subject to increased individual variation, probably related to their lack of experience in decoding cues for accent type based on the visual modality in isolation. Despite this, several native participants did discriminate accents through the visual signal with an accuracy of 70% - 90% correct, indicating that sufficient linguistic experience with a language enables the observer to decode visual cues relating to accent with reasonable accuracy.

The effect of regional accent type upon talker speechreadability was considered in Experiments 4A and 5 within this chapter, and also featured in Experiments 7, 8, 9 and 10 within Chapter 5, as illustrated by Figure 6.2. In each experiment, the Glaswegian talkers were found to be consistently more difficult to speechread than the Nottingham talkers. These results revealed the importance of accent type as a talker characteristic that directly influences talker speechreadability. Experiment 5 also enabled the question of observer expectations (i.e. observer accent type and consequent pattern of accent familiarity) upon speechreading performance to be analysed. The key finding was a significant interaction between participant location (Glasgow, Nottingham or Southampton) and accent type, which revealed a small, but significant, influence of accent familiarity on the Glaswegian participant's performance.

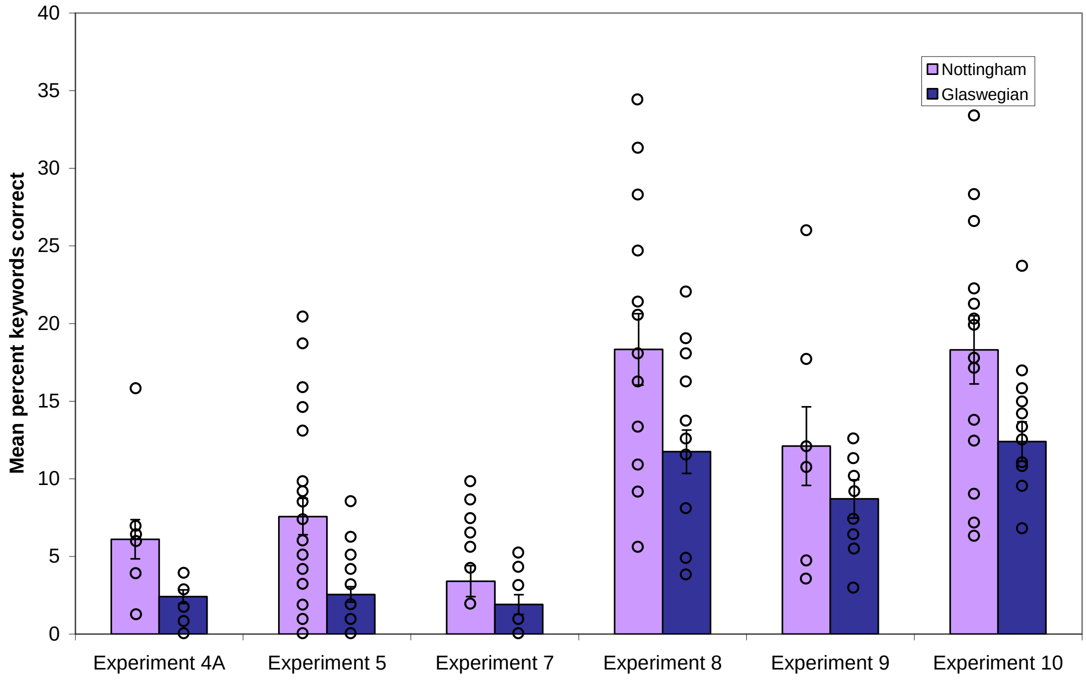


Figure 6.2: Mean percent keywords correct for Glaswegian and Nottingham talkers for Nottingham participants across six Experiments from Chapters 3 (4A and 5) and 5 (7, 8, 9, 10) shown with standard error bars. The spread of individual scores are plotted as dots for each Experiment.

The spread of individual scores for each Experiment indicates the high level of individual variability in speechreading performance. These relate to observer-specific factors which dictate the degree of accuracy with which an observer can decode visual cues and compensate for accent change.

6.2.3 Chapter 4: Accent type and talker variability

The study of talker and accent variability in Experiment 6 used an altered version of the Yakel and colleagues (2000) study to examine the influence of accent change on speechreading performance. Participants were asked to speechread sentences produced by multiple talkers with the same regional accent type, multiple talkers with different regional accents or a single talker. The study showed a decrement in speechreading performance associated with multiple-talker sentences lists when compared to single-talker lists. However, trial-by-trial variation of accent did not have an additive effect on talker variability effects. The lack of a positive result provides no support for the theory posited by Sommers and colleagues (1994) that the effects of talker variability upon speech processing are dependent on the number of levels across which talkers differ. Instead the theory put forward by Diehl and colleagues (1980) is more appropriate. The authors state that speech processing is 'reset' after each talker change. This negates the influence of talker-specific characteristics but produces a decrement in performance due to the utilisation of cognitive resources to process general talker change.

6.2.4 Chapter 5: reduction of accent effects through exposure, contextual constraints and repetition

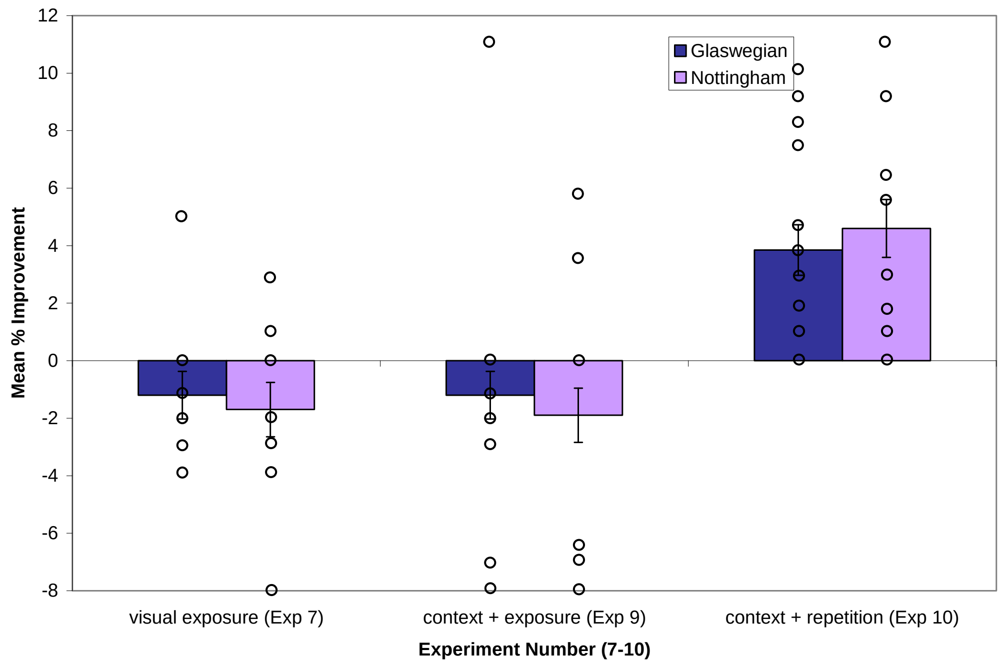


Figure 6.3: Mean percent improvement in performance for talkers with a Glaswegian or Nottingham accent in Experiments 7, 9 and 10, shown with standard error bars. The dots represent individual participant scores.

A series of four experiments analysed the effect of continued exposure, contextual cues and repetition upon accent effects on speechreading performance. Figure 6.3 illustrates a summary of the main results across three of the experiments, expressed as the mean difference in performance before and after application of the chosen method of reducing accent effects.

Interestingly in both cases of exposure, performance was actually *reduced* across talkers for the majority of the participants. The spread of individual performance was indicative of each participant's ability to extract linguistic information from the visual signal and to utilise contextual constraints. The large amount of variance in performance was expected based on past speechreading studies (Summerfield, 1992). Overall there were two key findings. First, it was apparent from the results that exposure to, and familiarisation with, the Glaswegian accent was insufficient to reduce accent effects on performance for the majority of the participants. This is indicative of a dichotomy in the processing of talker indexical information in the auditory and visual modalities. Second, the ability to utilise contextual cues has a greater influence on performance than an observer's ability to encode the articulatory cues provided by a talker. All participants' performance was facilitated by context and repetition

6.3 Discussion

6.3.1 Accent as a salient aspect of the visual signal

A talker's accent type influences many aspects of their speech production. These influences are portrayed through systematic, realisational, lexical-incidental and phonotactic differences (Wells, 1982a). There are two main categories of accent, both examined within this thesis; foreign and regional accent. Non-native, or foreign, productions of speech are thought to influence the production of speech to a greater extent than regional accent. This is because foreign-accented speech encompasses variation that is irrelevant to a native speaker, generated by phonetic and prosodic alterations to the speech signal that arise from their native language (Floccia et al, 2006). Regional accent variation, in comparison, represents recognisable articulatory variation within the confines of the language being spoken. Auditory research indicates that accent type is a salient aspect of the acoustic speech signal, with listeners able to identify an accent type from speech segments ranging in size from a sentence, to a syllable (Flege, 1984).

The level of information that can be extracted from the visual signal is somewhat ambiguous. Although there are strong visual correlates of speech relating to place of articulation (Summerfield, 1991), manner of articulation and voicing cues are much less visible (Summerfield, 1987). Consequently, comprehension of the visual signal is often far below that of auditory speech comprehension. Combined with evidence that points to the production of visemes being composed of several phonemes, resulting in the majority of words being relatively hard to distinguish using the visual signal alone (Owens & Blazek, 1985), some researchers suggest that the visual signal is not as information rich as the acoustic aspect of speech. However, more recent research suggests that the visual signal carries both adequate linguistic information for speech comprehension (Auer & Bernstein, 1997) and supplementary prosodic and indexical information. For example, information relating to lexical stress (Tye-Murray & Folkins, 1990) has been shown to be extracted from the visual signal. Furthermore, indexical information such as gender, age and emotional state can also be determined from a talker's appearance and facial cues (Belin et al, 2000).

The results of Experiments 1 and 4 consolidate more recent literature on visual speech by proposing that accent information can be extracted from the visual signal with some degree of success. Thus, the acoustic characteristics of an accent type influence the appearance of the visual signal, producing characteristic visual cues that an observer is able to decode in order to discriminate between accent types. This indicates that accent type is a salient aspect of the visual speech modality. This finding was present irrespective of the type of accent variation used as both foreign and regional accents were discriminated at a level above chance by all of the native English speaking participants. This finding illustrates the influence of accent upon the appearance of the visual signal and leads us to the first main finding of this thesis; the impact of accent type on talker speechreadability.

6.3.2: Talker speechreadability and accent type

Comprehensive literature from the auditory speech domain reports that it is accent unfamiliarity rather than the acoustic characteristics of an accent type that influences auditory speech processing (Anderson-Hsieh & Koehler, 1988; Floccia et al., 2006). Essentially, an unfamiliar accent alters the sound of speech, a listener is therefore unable to process the acoustic signal accurately in order to gain appropriate access to their internal lexicon and categorise the speech sounds. This mismatch between the perceived speech signal and the listener's stored representations of speech results in both slowed speech processing (Floccia et al, 2006) and increased misperceptions (Labov, 1989). This means that for every observer there is a specific pattern of accent familiarity that shapes their perception of speech. Thus a talker's accent characteristics may mould his or her production of speech, but it is the observer's *perception* of those characteristics, based on their home accent and contingent accent familiarity patterns, that denote talker intelligibility rather than the acoustic characteristics of the speech signal.

Research indicates, through examination of audiovisual speech, that information from both modalities is integrated at an early stage of speech processing (Arnold & Hill, 2001; McGurk & MacDonald, 1976). Hence, it was assumed that accent would exert a similar effect on visual speech processing to that found within auditory research. However, the second key finding of this thesis was that the visual correlates of accent type have a direct influence on talker speechreadability. Furthermore, this effect was not greatly reduced by observer-based patterns of familiarity. Based on the present results, the interpretation of the visual signal appears to be constrained to a greater degree by the characteristics of a talker's speech production than by the observer's generated expectations of visible articulations. Thus, the ability of an observer to extract the linguistic meaning of a message is defined by the visibility of a talker's articulations. The relative influence of an observer's home accent and consequent generation of expectations is relatively unclear, but appears to be a lesser influence than talker-based accent characteristics.

The finding that accent has a direct influence on the production of visual speech means that accent type represents a parameter of talker speechreadability. As such, the accent type of a talker should be considered in the generation of both tests of speechreading ability and training programmes designed to improve speechreading performance. Bench and colleagues (1995) state that in order for a test of speechreading to be valid, a range of talkers should be selected. The results detailed here suggest that accent type should therefore be a consideration when selecting talkers for a test of speechreading. A range of accent types should be included if the test is one of general speechreading ability. Furthermore, when training an observer in speechreading it would be advantageous to include a variety of accent types in the training programme in order to provide a general overview of the type of variability an observer will encounter in an everyday environment. Finally, certain 'difficult' accent types, such as the Glaswegian accent, should perhaps be excluded from tests of speechreading in order to avoid the floor effects found in several studies within this thesis.

6.3.3 Accent 'normalisation' and theories of speech processing

There are two main theories of auditory speech processing; 'abstractionist' and 'encoding'. Abstractionist theory regards talker variation as an extraneous source of 'noise' which must be removed from the acoustic signal (Nygaard & Pisoni, 1998). Consequently indexical information is removed from the speech signal at an early stage of processing. Lexical matching of the acoustic signal to the listener's internal lexicon is therefore conducted on the basis of a pre-normalised and abstract signal. This theory posits that speech comprehension will be initially disrupted by an unfamiliar or highly variable signal. However, after a period of adaptation, comprehension levels should return to the previous level. In comparison encoding theories view talker-specific characteristics as valid aspects of the speech signal. Successful perception is therefore achieved through the encoding of indexical characteristics such as accent into the internal lexicon (Nygaard & Pisoni, 1998). However, a consequence of retaining talker characteristics is that comprehension will be reduced when a listener must encode multiple sets of talker information. Again, after a period of adjustment, in which talker characteristics are encoded into memory, comprehension should return to baseline. In terms of accent processing, research suggests that the encoding of unfamiliar accent characteristics and subsequent return to previous levels of performance is swift, with as little as one minute required for full adaptation to take place (Clarke & Garrett, 2004).

The application of these theories to visual speech would indicate that i) speechreading performance should be adversely affected by variation in the visual signal caused by accent, ii) the use of multiple talkers should have a detrimental affect on speechreading performance when compared to a single talker, iii) after a period of adaptation, the perception of a talker with an unfamiliar accent should return to baseline.

I have already detailed in the previous section that accent type has a direct effect on talker speechreadability and thus speechreading performance. The next example to be considered was the influence of multiple talkers upon visual speech perception. Chapter 4 consolidates previous research by Yakel and colleagues (2000) by illustrating the detrimental influence of multiple talkers upon speechreading performance. This finding suggested that, similar to auditory speech processing, variation in the visual signal must be normalised or encoded in order to facilitate the perception of multiple talkers. This finding indicates a similarity in visual and auditory speech processing but does not provide strong evidence for either theory of speech perception.

Further analysis incorporated the results of Experiments 7, 8, 9 and 10 which indicated that, unlike similar auditory research, exposure alone was not sufficient to reduce the effect of accent on visual speech perception. It therefore seems likely that the processing strategies utilised by observers when faced with talker variation in the visual signal differ from those utilised by listeners presented with acoustic talker variation. Certainly, exposure to the Glaswegian accent alone was insufficient to improve an observer's ability to process the visual signal. Thus, the results do not provide any evidence for the encoding theory of speech perception. However, a potential explanation of the results is that observers were attempting to remove the variation from the visual signal caused by multiple talkers and the Glaswegian accent in order to 'normalise' the signal. In so doing, they utilised cognitive resources and were consequently less able to process the visual signal, producing the recorded decrement in performance after training in Experiments 7 and 9. In summary, the research detailed here provides only partial evidence for the applicability of either theory of speech perception to visual speech processing. Further research in this area is therefore required.

6.3.4 Observer factors in speechreading performance

Thus far it would appear the characteristics of a talker have a greater influence on performance than observer factors. However, the spread of individual performance shown across all of the experiments suggest that observer-factors also exert an influence on performance. There are several theories which attempt to explain this wide variance in individual performance, which is much larger than that found in tests of auditory processing ability (Summerfield, 1992). First, speechreading ability does not appear to correlate with most measures of cognitive abilities. Second, speechreading performance has been found to correlate with a latent response to visually evoked scalp potentials (bright light) (Shepherd, DeLavergne, Frueh & Clobridge, 1977). This has led some researchers to suggest that speechreading ability is related to low-level neural processing and as such represents a physiologically based skill, explaining some of the variability in performance (see Summerfield, 1991; 1992 for a review). Alternatively, more recent research points to large working memory capacity and a high speed of lexical processing as neural correlates of proficient speechreading performance (Ronnberg, Andersson, Samuelsson, Soderfeldt, Lyxell & Risberg, 1999). It is possible that some of the capacity to deal with accent variation is related to these aspects of speechreading ability. Certainly, a general finding throughout the research was that proficient speechreaders were able to speechread the Glaswegian talkers more accurately than the poor speechreaders, though the advantage for the Nottingham accent always remained. This may be directly related to their working memory capacity and visual processing skills.

However, of particular interest within this thesis was the ability of observers to utilise contextual constraints and repair strategies such as repetition. Grant and Seitz (2000) suggest that an observer's ability to speechread is dictated by their ability to i) extract linguistic cues from the visual signal and ii) utilise lexical constraints to minimise the linguistic possibilities of an articulatory pattern. The results of Chapter 5 indicate that the ability to utilise lexical constraints could have a direct influence on both speechreading performance and an observer's ability to speechread the Glaswegian accent. Certainly the spread of scores illustrated by Figures 6.2 and 6.3 indicate that not all observers benefited equally from the provision of a contextual cue. This result has a wider implication as the ability to utilise contextual constraints is tied to memory capacity, vocabulary, and the skill of inference-making (Grant & Seitz, 2000), all of which could potentially influence an observer's ability to speechread and to compensate for variation in the visual signal. This is a consideration for future speechreading training programmes as the inclusion of contextual cues may be a useful tool if accent or talker variation constitutes part of the programme. In summary, whilst it is apparent that an observer's pattern of accent familiarity may have only a small influence on speechreading performance, it is likely that an observer's viewing strategy will have an influence. There are many observer factors that might therefore influence the ability of that observer to process accented speech. Unfortunately the parameters of those abilities were not covered by this thesis and therefore remain a question for future research.

6.4 Conclusion

The characteristics of an accent type represent a visually salient aspect of talker indexical information. The visual correlates of accent type have a direct influence on the clarity of the visual signal produced and thus represent a parameter of talker speechreadability. Accent therefore has an effect on speechreading performance that is resilient, replicable and difficult to reduce. Although observer-related accent familiarity patterns appear to have little influence upon speechreading performance, other factors such as the ability to utilise contextual constraint do have an influence. Overall, the results of this thesis suggest that accent should be an important consideration for the development of any future tests of speechreading ability or training programmes.

6.5 Further Research

The research presented here presents a strong case for the importance of accent type in visual speech processing. However, several questions remain which would benefit from further research in this area.

First, while I have shown that observers were able to extract cues relating to accent from the visual signal, the individual variability in performance was not explained. Research into observer viewing strategy has indicated that observers differ in their method of extracting linguistic cues from the visual signal (Conrey

& Gold, 2006). The authors infer from eye-movement data, that observers differ in their focus point of the talker's face. Some observers viewed the entire face, whilst others focused on the talker's mouth. It is likely that the participants used in Experiments 1 and 4 would have shown similar differences in their viewing strategies, perhaps explaining some of the noted individual variability. Furthermore research into accent discrimination utilising similar eye-tracking methodology to that used by Conrey and Gold (2006) could determine those aspects of facial motion which carry the most relevant accent cues. This would help to define those aspects of accent which have the strongest visual correlates and are therefore the most important indicators of accent type.

Second, the British accent effects analysed throughout this thesis relate to accent differences on a regional level, as prescribed by the five main accent regions within the UK (Wells, 1982b). The results indicate that the visual processing system is sensitive to broad accent changes. However, it may be the case that the Glaswegian accent is a special case of accent type, one which is universally difficult to process. Further research using different accent types would be required to determine if similar effects are found across different regional accents. Furthermore, research utilising accents that differ on a sub-regional level (such as a comparison of Sheffield and Nottingham accents) could help to determine the sensitivity of the visual processing system to accents which share a greater number of characteristics. This is an important consideration if accent is to be considered as a parameter of talker speechreadability. Thus, the degree of difference necessary between two accents before an effect is seen should be quantified.

Finally, the influence of observer factors upon speechreading performance across accent types has not been thoroughly quantified within this thesis. Further research utilising measures of observer variability could illuminate the factors that are necessary for both efficient utilisation of contextual constraints and compensation of accent effects. This could have a direct influence on the production of training programmes designed to include accent as an aspect of the visual signal.

References

Abercrombie, D. (1967) *Elements of General Phonetics,* Chicago: Aldine

Adank, P. & McQueen, J.M. (2007) The effect of an unfamiliar regional accent on spoken word comprehension, *Proceedings of the XVIth International Congress of Phonetic Sciences,* 1387-1392

Anderson-Hsieh, J. & Koehler, K. (1988). The effect of foreign accent and speaking rate on native speaker comprehension. *Language learning, 38, 4,* 561-613

Arnold, P. (1997) The structure and optimization of speechreading, *The Journal of Deaf Studies and Deaf Education, 2,* 199 - 211

Arnold, P. & Hill, F. (2001) Bisensory augmentation: a speechreading advantage when speech is clearly audible and intact, *British Journal of Psychology, 92,* 339-355

Assman, P.F. Nearey, T.M. & Hogan, J.T. (1982) Vowel identification: Orthographic, perceptual and acoustic aspects, *The Journal of the Acoustical Society of America, 71, 4,* 975-989

Auer, E.T. Bernstein, L.E. (1997) Speechreading and the structure of the lexicon: Computationally modeling the effects of reduced phonetic distinctiveness on lexical uniqueness, *Journal of the Acoustical Society of America, 102, 6,* 3704-3710

Bardhun, N.P. Aslin, R.N. & Tanenhaus, M.K. (2006) Return of the Weckud Wetch: Rapid Adaptation to a new accent. *Journal of the Acoustical Society of America, 119, 5,* 3423

Belin, P. Fecteau, S. & Bedard, C. (2004) Thinking the voice: neural correlates of voice perception, *TRENDS in Cognitive Science, 8, 3,* 129-135

Bench, J. Kowal, A. & Bamford, J. (1979). The BKB (Bamford-Kowal-Bench) sentence lists for partially hearing children. *British Journal of Audiology, 13,* 108-112

Bench, J. Daly, N. Doyle, J. & Lind, C. (1995) Choosing talkers for the BKB/A test: A procedure with observations on talker age and gender, *British Journal of Audiology, 29, 3,* 172-187

Bent, T. & Bradlow, A.R. (2003) The interlanguage speech intelligibility benefit. *Journal of the Acoustical Society of America, 114, 3,* 1600-1610

Berger, K.W. (1972) *Speechreading: Principles and Methods,* National Educational Press, Maryland, USA, Chapters 5 & 6, 109-150

Bernstein, L.E. Auer, E.T. & Tucker, P.E. (1998) *Does training enhance visual speech perception?* House Ear Institute, California

Bernstein, L.E. Auer, E.T. & Tucker, P.E. (2001) Enhanced speechreading in deaf adults: can short-term training/practice close the gap for hearing adults. *Journal of Speech, Language and Hearing Research, 44,* 5-18

Bertelson, P. Vroomen, J. & Gelder, B. (2003) Visual recalibration of auditory speech identification: a McGurk after effect, *Psychological Science, 14, 6,* 592-597

Bodner, G.E. & Masson, M.E.J. (1997) Masked repetition priming of words and nonwords: Evidence for nonlexical basis for priming, *Journal of Memory and Language,* 37, 268 – 293

Boothroyd, A. (1988) Linguistic factors in speechreading, in, C.L. De Filippo & D.G. Sims (eds) *The Volta Review, New Reflections on Speechreading, 90, 5,* 77-87

Bradlow, A.R & Bent, T. (2003) Listener Adaptation to foreign-accented English. M.J. Sole, Recasens, D. & J. Romero (eds) *Proceedings of the XVth International Congress of Phonetic Sciences,* 2881-2884

Bruce, V. Young, A. (1986) Understanding face recognition. *The British Journal of Psychology, 77, 3,* 305-327

Campbell, R. Brooks, B. Deflann, E. & Roberts, T. (1996) Dissociating face processing skills: Decisions about lip-speech, expression and identity, *The Quarterly Journal of Experimental Psychology, 49A, 2,* 295-314

Campbell, R. Landis, T. & Regard, M. (1986) Face recognition and lipreading a neurological dissociation, *Brain, 109,* 509-521

Calabresi, M. (2007) Early visual and sentence level contributions to speechreading, *unpublished PhD thesis*

Carr, P. (1999) *English Phonetics and Phonology, An Introduction,* Blackwell Publishing, Oxford

Clarke, C.M. & Garrett, M.F. (2004) Rapid adaptation to foreign-accented English, *Journal of the Acoustical Society of America, 116, 6,* 3647-3658

Clopper, C.G. & Bradlow, A.R. (2006) Effects of dialect variation on speech intelligibility in noise, *Journal of the Acoustical Society of America, 119, 5,* 3424

Cohen, L.H. (1995) *Train Go Sorry; Inside a Deaf World,* Vintage Books, New York, Chapter 12, 195

Cohen, M.M. & Massaro, D.W. (1993) Modelling co-articulation in synthetic visual speech, In N. Magnenat-Thalmann & D. Thanlmann (eds) *Models and Techniques in Computer Animation,* 139-156

Cole, R.A. Coltheart, M. & Allard, F. (1974) Memory of a speaker's voice: Reaction time to same- or different-voiced letters, *The Quarterly Journal of Experimental Psychology, 26,* 1-7

Cole, R.A. Jakimik, J. & Cooper, W.E. (1980) Segmenting speech into words, *Journal of the Acoustical Society of America, 67, 4,* 1323-1332

Colin, C. Radeau, M. Soquet, A. Demolin, D. Colin, F. & Deltenre, P. (2002) Mismatch negativity evoked by the McGurk-MacDonald effect: a phonetic representation within short term memory, *Clinical Neurophysiology, 113,* 495-506

Connolly, J.F. & Phillips, N.A. (1994) Event-related potential components reflect phonological and semantic processing of the terminal word of spoken sentences, *Journal of Cognitive Neuroscience, 6, 3,* 256 - 266

Conrey, B. & Gold, J.M. (2006) An ideal observer analysis of variability in visual-only speech, *Vision Research, 46, 19,* 3243-3258

Cox, R.M. Alexander, G.C. & Gilmore, C. (1987) Intelligibility of average talkers in typical listening environments, *The Journal of the Acoustical Society of America, 81, 5,* 1598-1608

Davis, C. & Kim, J. (1998). Repeating and remembering foreign language words: does seeing help? *Proceedings of Auditory – Visual Speech Processing 1998*

Davis, C. & Kim, J. (2001). Repeating and remembering foreign language words: Implications for language teaching systems. *Artificial Intelligence Review, 16,* 37-47

Dennis, I. & Schmidt, K. (2003) Associative processes in repetition priming, *Journal of Experimental Psychology: Learning, Memory and Cognition, 29, 4,* 532-538

Demorest, M.E. & Bernstein, L.E. (1992) Sources of variability in speechreading sentences: a generalizability analysis. *Journal of Speech and Hearing Research, 35,* 876 - 891

Diehl, R.L. Souther, A.F. & Convis, C.L. (1980) Conditions of rate normalisation in speech perception, *Perception and Psychophysics, 27,* 435-443

Ellis, H.D. (1986) Processes underlying face recognition, *In,* R.Bruyer (ed), *The neuropsychology of face perception and facial expression,* Hillsdale, Hew Jersey, Lawrence Erlbaum Associates

Ellis, T. MacSweeney, M. Dodd, B. & Campbell, R. (2001) TAS: A new test of adult speechreading. Deaf people really can be better speechreaders. *Proceedings from AVSP 2001 International Conference on Auditory-Visual Speech Processing*

Erber, N. (1992) Effects of question-answer format on visual perception of sentences, *Journal of the Academy of Rehabilitative Audiology, 25,* 113-122

Evans, B.G. & Iverson, P. (2003) Vowel normalisation for accent: An investigation of the best exemplar locations in northern and southern British English sentences, *The Journal of the Acoustical Society of America, 115, 1,* 352-361

Fitzer, J. (2003) Speechreading – Not just for deafies, *Journal of the Association of Medical Professionals with Hearing Losses, 1, 2,* 1-4

Flege, J.E. (1984) The detection of French accent by American listeners, *The Journal of the Acoustical Society of America, 76, 3,* 692-707

Flege, J.E. (1992) The intelligibility of English vowels spoken by British and Dutch talkers. R.D. Kent. (ed) *Intelligibility in Speech Disorders, Chapter 5,* 157-232

Flege, J.E. & Hillenbrand, J. (1984) Limits on phonetic accuracy in foreign language speech production, *Journal of Acoustical Society of America, 76, 3,* 708-721

Floccia, C. Girard, F. Goslin, J. & Konopczynski (2006) Does regional accent perturb speech processing? *Journal of Experimental Psychology: Human Perception and Performance, 32, 5,* 1276-1293

Flynn, M.C. & Dowell, R.C. (1999) Speech perception in a communicative context: An investigation using question/answer pairs. *Journal of Speech, Language and Hearing Research, 42,* 540-552

Gagne, J.P. Charest, M. Le Monday, K. & Desbiens, C. (2006). Evaluation of an audio-visual FM system: Speechreading performance as a function of distance. *International Journal of Audiology, 45, 5,* 295-300

Gass, S. & Varonis, E.M. (1984) The effect of familiarity on the comprehensibility of non-native speech, *Language Learning, 34, 1,* 65-87

Gesi, A.T. Massaro, D.W. Cohen, M.M. (1992) Discovery and expository methods in teaching visual consonant and word identification, *Journal of Speech and Hearing Research, 35,* 1180-1188

Goldinger, S.D., Pisoni, D.B. & Logan, J.S. (1991) On the locus of talker variability effects in recall of spoken word lists, *Journal of Experimental Psychology: Learning, Memory and Cognition, 17,* 152 - 162

Grant, K.W. & Seitz, P.F. (2000) The recognition of isolated words and words in sentences: Individual variability in the use of sentence context. *Journal of the Acoustical Society of America, 107, 2,* 1000 - 1011

Green, K.P. Kuhl, P.K. Meltzoff, A.N. Stevens, E.B. (1991) Integrating speech information across talkers, gender and sensory modality: Female faces and male voices in the McGurk effect. *Perception and Psychophysics, 50, 6,* 524-536

Hazan, V. Sennema, A. Faulkner, A. Ortega-Llebaria, M. Iba, M. & Chung, H. (2006) The use of visual cues in the perception of non-native consonant contrasts, *Journal of the Acoustical Society of America, 119, 3,* 1740 - 1751

Helfer, K.S. & Freyman, R.L. (2005) The role of visual speech cues in reducing energetic and informational masking. *Journal of the Acoustical Society of America, 112, 2,* 842 - 849

Howell, P. Barry, W. & Vinson, D. (2006) Strength of British accents in altered listening conditions, *Perception and Psychophysics, 68, 1,* 139-153

Hughes, A. & Trudgill, P. (1987) *English Accents and Dialects, An introduction to Social and Regional Varieties of British English,* Chapter 1, Edward Arnold, London

Hughes, A. Trudgill, P. & Watt, D. (2005) *English Accents and Dialects, An Introduction to Social and Regional Varieties of English in the British Isles (4th Edition),* Hodder Education, London

Ikeno, A. & Hansen, J.H.L. (2006) Perceptual recognition cues in native English accent variation: 'Listener accent, perceived accent and comprehension', *Acoustics, Speech and Signal Processing IEEE 2006, 1,* I401-I404

Ikeno, A. & Hansen, J.H.L. (2007) The effect of listener accent background on accent perception and comprehension, *EURASIP Journal on Audio, Speech and Music Processing 2007,* 1-8

Ijsseldijk, F.J. (1992) Speechreading performance under different conditions of video image, repetition and speech rate, *Journal of Speech and Hearing Research, 35,* 466 – 471

Jacobs, M.A. (1982) Visual communication (speechreading) for the severely and profoundly hearing-impaired young adult, *In*, D.G. Sims, G.G. Walter & R.L. Whitehead (eds), *Deafness and communication: Assessment and training,* Baltimore, M.D. Williams and Wilkins

Jeffers, J. & Barley, M. (1971) *Speechreading (lipreading),* Thomas, Springfield, IL:C.C.

Jiang, J. Alwan, A. Keating, P.A. Auer, E.T. & Bernstein, L.E. (2002) On the relationship between face movements, tongue movements and speech acoustics. *Journal of Applied Signal Processing*, *11,* 1174 - 1188

Jordan, T.R. McCotter, M.V. & Thomas, S. (2000) Visual and audiovisual speech perception with colour and gray scale facial images. *Perception and Psychophysics, 62, 7,* 1394-1404

Jordan, T.R. & Sergeant, P. (2000) Effects of distance on visual and audiovisual speech recognition, *Language and Speech, 43,* 107-124

Jordan, T.R. & Thomas, S. (2001) Effects of horizontal viewing angle on visual and audiovisual speech recognition, *Journal of Experimental Psychology: Human Perception and Performance, 27, 6,* 1386-1403

Kaiser, A.R. Kirk, K.I. Lachs, L. & Pisoni, D.B. (2003) Talker and lexical effects on audiovisual word recognition by adults with cochlear implants. *Journal of Speech, Language and Hearing Research, 46,* 390 - 404

Kamachi, M. Hill, H. Lander, K. & Vatikiotis-Bateson, E. (2003) 'Putting the face to the voice' matching identity across modality. *Current Biology, 13, 19,* 1709-1714

Kim, J. Davis, C. & Krins, P. (2004) Amodal processing of visual speech as revealed by priming, *Cognition, 93,* B39 – B47

Kraljic, T. & Samuel, A.G. (2005) Perceptual Learning for speech: Is there a return to normal? *Cognitive psychology, 51,* 141-178

Kraljic, T. & Samuel, A.G. (2007) Perceptual adjustments to multiple speakers. *Journal of Memory and Language, 56, 1,* 1-15

Kricos, P.B. & Lesner, A. (1982) Differences in visual intelligibility across talkers, *The Volta Review, 84,* 219 – 225

Kricos, P.B. & Lesner, A. (1985) Effect of talker differences on the speechreading of hearing impaired teenagers. *The Volta Review, 87,* 5-16

Kutas, M. & Hillyard, S.A. (1980) Reading senseless sentences: Brain potentials reflect semantic incongruity, *Science, 207,* 203-204

Labov, W. (1989) The Limitations of Context: Evidence from misunderstandings in Chicago, *Chicago Linguistic Society, 25,* 171-200

Labov, W. & Ash, S. (1997) Understanding Birmingham, in C. Bernsteing, T. Nunnally & R. Sabino (eds) *Language Variety in the South Revisited,* Alabama University, 508 - 573

Lander, K. & Davies, R. (2008) Does face familiarity influence speechreadability? *The Quarterly Journal of Experimental Psychology, 61, 7,* 961-967

Lansing, C.R. & Helgeson, C.L. (1995) Priming the visual recognition of spoken words. *Journal of Speech and Hearing Research, 38,* 1377-1386

Lesner, S.A. (1988), The Talker, *In* C.L. Fillipo & D.G. Sims (eds) *New Reflections on Speechreading: The Volta Review*, 90, 5, 89-98

Lesner, S.A & Kricos, P.B. (1981) Visual vowel and dipthong perception across speakers, *Journal of the Academy of Rehabilitative Audiology, 14,* 252-258

Levi, S.V. & Pisoni, D.B. (2007) Indexical and linguistic channels in speech perception: Some effects of voiceovers on advertising outcomes, in, T.M.Lowrey (ed) *Psycholinguistic Phenomena in Marketing Communications,* 203-219

Lidestam, B. Lyxell, B. & Lundeberg, M. (2001) Speechreading of synthetic and natural faces: Effects of contextual cueing and mode of presentation, *Scandinavian Audiology, 30, 2,* 89-94

Lisker, L. & Rossi, M. (1992) Auditory and visual cueing of the (± rounded) feature of vowels, *Language and Speech, 35,* 391-417

MacLeod, A. & Summerfield, Q. (1987) Quantifying the contribution of vision to speech perception in noise, *British Journal of Audiology, 21, 22,* 131-141

Magen, H.S. (1998). The perception of foreign-accented speech, *Journal of Phonetics, 26,* 381-400

Martin, C.S. Mullenix, J.W. Pisoni, D.B & Summers, W.V. (1989) Effects of talker variability on recall of spoken word lists, *Journal of Experimental Psychology: Learning Memory and Cognition, 15,4 ,* 676-684

Marzolf, C.A. Stewart, M. Nerbonne, M.A. & Lehman, L.E. (1998) Effects of two repair strategies on speechreading of words and sentences, *Journal American Academy Audiology, 9,* 243-248

Massaro, D.W. Cohen, M.M. & Gesi, A.T. (1993) Long-term training, transfer and retention in learning to lipread, *Perception and Psychophysics, 5,* 549-562

Massaro, D.W. & Light, J. (2004) Using visible speech to train perception and production of speech for individuals with hearing loss, *Journal of Speech, Language and Hearing Research, 47,* 304 -320

Mattys, S.L. (2000) The perception of primary and secondary stress in English, *Perception and Psychophysics, 6, 2,* 253-265

Maye, J. Aslin, R.N. & Tanenhaus, M.K. (2008) The weckud wetch of the wast: Lexical adaptation to a novel accent, *Cognitive Science, 32,* 543-562

McGurk, H & MacDonald, J. (1976) Hearing lips and seeing voices, *Nature, 264,* 746-748

Miller, G.A. Heise, G.A. & Lichten, W. (1951) The intelligibility of speech as a function of the context of the test materials, *Journal of Experimental Psychology, 41, 5,* 329-335

Mills, A.E. (1987) The development of phonology in the bind child, *in,* Dodd, B & Campbell, R. (eds) (1987) *Hearing By Eye: The Psychology of Lip-Reading,* Lawrence Erlbaum Associates, London

Montgomery, A.A. & Jackson, P.L. (1983) Physical characteristics of the lips underlying vowel lipreading performance, *The Journal of the Acoustical Society of America, 73, 6,* 2134-2144

Most, T. & Adi-Bensaid, L. (2001) The influence of contextual information on the perception of speech by post-lingually and pre-lingually profoundly hearing-impaired Hebrew speaking adolescents and adults, *Ear and Hearing, 22, 3,* 252-263

Mullenix, J.W. Pisoni, D.B. & Martin, C.S. (1989) Some effects of talker variability on spoken word recognition, *Journal of the Acoustical Society of America, 85, 1,* 365 - 378

Munhall, K.G. Jones, J.A. Callan, D. Kuratate, T. & Vatikiotis-Bateson, E. (2003) Visual prosody and speech intelligibility: Head movement improves auditory speech perception, *Psychological Science, 15, 2,* 133-137

Munro, M.J. & Derwing, T.G. (1995). Processing time, accent and comprehensibility in perception of native and foreign-accented speech. *Language and Speech, 38,* 289-306

Nathan, L. Wells, B. & Donlan, C. (1998) Childrens comprehension of unfamiliar regional accents: a preliminary investigation, *Journal of Child Language, 25,* 343-365

Nathan, L. & Wells, B. (2001) Can children with speech difficulties process an unfamiliar accent? *Applied Psycholinguistics, 22,* 343-361

Navarra, J. & Soto-Faraco, S. (2007) Hearing lips in a second language: visual articulatory information enables the perception of second language sounds, *Psychological Research, 71*, 4 - 12

Neri, A. Cucchiarini, C. & Strik, H. (2006) Selecting segmental errors in non-native Dutch for optimal pronunciation training, *Internationl Review of Applied Linguistics, 44,* 367 – 404

Norris, D. McQueen, J.M. & Cutler, A. (2003) Perceptual learning in speech, *Cognitive Psychology, 47, 2,* 204-238

Nygaard, L.C., Sommers, M.S. & Pisoni, D.B. (1994) Speech Perception as a talker contingent process, *Psychological Science, 5,* 42 - 46

Nygaard, L.C. & Pisoni, D.B. (1998) Talker-specific learning in speech perception, *Perception and Psychophysics, 60, 3,* 355 – 376

Owens, E. & Blazek, B. (1985) Visemes observed by hearing-impaired and normal-hearing adult viewers, *Journal of Speech, Language and Hearing Research, 28,* 381-393

Palmer, L. (1988) Speechreading as communication, *The Volta Review, 90, 5,* 33-44

Pisoni, D.B. (1997) Some thoughts on 'normalization' in speech perception, 9 -32 *in* Johnson, K. & Mullennix, J.W. (eds) (1997) *Talker Variability in Speech Processing,* Academic Press

Reisberg, D. McLean, J. & Goldfield, A, (1987) Easy to hear but hard to understand: a lipreading advantage with intact auditory stimuli, 97 - 113 *in* Dodd, B & Campbell, R. (eds) (1987) *Hearing By Eye: The Psychology of Lip-Reading,* Lawrence Erlbaum Associates, London

Rochet, B. (1995) Perception and production of L2 speech sounds, *In,* W. Strange (ed), *Perception and linguistic experience: theoretical and methodological issues,* Timonium, MD, York Press

Rogers, R.D & Monsell, S, (1995) Costs of a predictable switch between simple cognitive tasks, *Journal of Experimental Psychology, 124, 2,* 207-231

Ronnberg, J. Andersson, J. Samuelsson, S. Soderfeldt, B. Lyxell, B. & Risberg, J. (1999) A speechreading expert: The case of MM, *Journal of Speech, Language and Hearing Research, 42,* 5-20

Rosenblum, L.D. Miller, R.M. & Sanchez, K. (2007) Lip-read me now, hear me better later, cross modal transfer of talker familiarity effects, *Psychological Science, 18, 5,* 392-396

Rosenblum, L.D. Johnson, J. & Saldana, H.M. (1996) Point-light facial displays enhance comprehension of speech in noise, *Journal of Speech and Hearing Research, 39,* 1159-1170

Rosenblum, L.D. Schmuckler, M. & Johnson, J. (1997) The McGurk effect in infants, *Perception and Psychophysics, 59, 3,* 347 – 357.

Rosenblum, L.D. Smith, N.M. Nichols, S.M. Hale, S. & Lee, J. (2006) Hearing a face: Cross-modal matching using isolated visible speech, *Perception and Psychophysics, 68, 1,* 84 – 93.

Rosenblum, L.D. & Yakel, D.A. (2001) The McGurk effect from single and mixed speaker stimuli, *Acoustics Research Letters Online,* 67-72

Schmid, P.M. & Yeni-Komshian, G.H. (1999) The effects of speaker accent and target predictability on perception of mispronunciations, *Journal of Speech, Language and Hearing Research, 42,* 56-64

Schweinberger, S.R & Soukup, G.R. (1998) Asymmetric relationships among perceptions of facial identity, emotion and facial speech, *Journal of Experimental Psychology: Human Perception and Performance, 24, 6,* 1748-1765

Sheffert, S.M. Olsen, E. (2004) Audiovisual speech facilitates voice learning, *Perception and Psychophysics, 66, 2,* 352 - 362

Shepherd, D.C. DeLavergne, R.W. Frueh, F.X. & Clobridge, C. (1977) Visual-neural correlate of speechreading ability in normal-hearing adults, *Journal of Speech, Language and Hearing Research, 20, 4,* 752-765

Sommers, M.S. Nygaard, L.C. & Pisoni, D.B. (1994) Stimulus variability and spoken word recognition. I. Effects of variability in speaking rate and overall amplitude, *Journal of the Acoustical Society of America, 96, 3,* 1314-1324

Soto-Faraco, S., Navarra, J., Weikum, W.M., Vouloumanos, A., Sebastian-Galles, N. & Werker, J.F. (2007), Discriminating Languages by Speechreading, *Perception and Psychophysics, 29, 2,* 218-231

Squires, S.C. & Dancer, J.E. (1986) Auditory versus visual practice effects in the intelligibility of words in everyday sentences, *Journal of Auditory Research, 26,* 5-10

Stuart-Smith, J. (1999) Glasgow: accent and voice quality, *in*, Foulkes, P. & Docherty, G. (eds) (1999) *Urban Voices: Accent Studies in the British Isles, Chapter 11,* 203-222

Stevens, A.A. (2004) Dissociating the cortical basis of memory for voices, words and tones, *Cognitive Brain Research, 18,* 162-171

Sumby, W.H. & Pollack, I (1954) Visual contribution to speech intelligibility in noise, *The Journal of the Acoustical Society of America,* 26, 2, pp. 212-215

Summerfield, Q. (1987) Some preliminaries to a comprehensive account of audiovisual speech perception *in,* Dodd, B & Campbell, R. (eds) (1987) *Hearing By Eye: The Psychology of Lip-Reading,* Lawrence Erlbaum Associates, London

Summerfield, Q. (1991) Visual perception of phonetic gestures *in,* Mattingly, I.G. & Studdert-Kennedy, M. (eds) (1991) *Modularity and the Motor Theory of Speech*

Perception; Proceedings of a Conference to Honour Alvin M. Liberman, Lawrence Erlbaum Associates, London

Summerfield, Q. (1992) Lipreading and audio-visual speech perception, *Philosophical Transactions: Biological Sciences, 335,* 71-78

Summerfield, Q & McGrath, M. (1984) Detection and resolution of audio-visual incompatibility in the perception of vowels, *The Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology, 36A,* 51-74

Taylor, W.L. (1953) Cloze procedure: A new tool for measuring readability, *Journalism Quarterly, 30,* 415-433

Tranel, B. (1987) *The sounds of French: An introduction,* Cambridge University Press

Tye-Murray, N. Purdy, S.C. Woodworth, G.C. & Tyler, R.S. (1990) Effects of repair strategies on visual identification of sentences, *Journal of Speech and Hearing Disorders, 55,* 621-627

Tye-Murray, N. & Folkins, J.W. (1990) Jaw and lip movements of deaf talkers producing utterances with known stress patterns, *Journal of Acoustical Society of America*, *87, 6,* 2675-2683

Verbrugge, R.R. Strange, W. Shankweiler, D.P. & Edman, T.R. (1976) What information enables a listener to map a talkers vowel space? *Journal of the Acoustical Society of America, 60,* 198 - 212

Walden, B.E. Prosek, R.A. Montgomery, A.A. Scherr, C.K. & Jones, C.J. (1977) Effects of training on the visual recognition of consonants, *Journal of Speech and Hearing Research,* Vol. 20, 130-145

Walden, B.E. Erdman, S.E. Montgomery, A.A. Schwartz, D.M. & Prosek, R.A. (1981) Some effects of training on speech recognition by hearing impaired adults, *Journal of Speech and Hearing Research, 24,* 207-216

Walker, S. Bruce, V. O'Malley, C. (1995) Facial identity and facial speech processing: familiar faces and voices in the McGurk effect, *Perception and Psychophysics, 57, 8,* 1124 – 1133

Weikum, W.M. Vouloumanous, A. Navarra, J. Soto-Faraco, S. Sebastian-Galles, N. & Werker, J.F. (2007) Visual language discrimination in infancy, *Science, 316,* 1159

Wells, J.C. (1982a) *Accents of English 1: An Introduction,* University Press, Cambridge

Wells, J.C. (1982b) *Accents of English 2: The British Isles,* University Press, Cambridge

Wijngaarden, S.J.V, Steeneken, H.J.M, Houtgast, T. (2002) Quantifying the intelligibility of speech in noise for non-native listeners, *Journal of the Acoustical Society of America, 111, 4*, 1906-1916

Yakel, D.A. Rosenblum, L.D. & Fortier, M.A. (2000) Effects of talker variability on speechreading, *Perception and Psychophysics, 62, 7,* 1405-1412

Yonan, C.A. & Sommers, M.S. (2000) The effects of talker familiarity on spoken word identification in younger and older listeners, *Psychology and Ageing, 15, 1,* 88-99

Appendix 1

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 ɡ	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.

Table A.1 The International Phonetic Alphabet (revised to 2005, produced by the International Phonetic Association)

Vowel	Lexical set	Examples
ɪ	1. Kit	Ship, sick, bridge…
e	2. Dress	Step, neck, edge…
æ	3. Trap	Tap, back, badge…
ɒ	4. Lot	Stop, sock, dodge…
ʌ	5. Strut	Cup, suck, pulse…
ʊ	6. Foot	Put, bush, full…
ɑ ː	7. Bath	Staff, brass, ask…
ɒ	8. Cloth	Cough, broth, cross…
ɜ ː	9. Nurse	Hurt, lurk, burst…
i ː	10. Fleece	Creep, speak, leave…
e ɪ	11. Face	Tape, cake, leave…
ɑ ː	12. Palm	Psalm, father, bra…
ɔ ː	13. Thought	Taught, sauce, hawk…
ə ʊ	14. Goat	Soap, joke, home…
u ː	15. Goose	Loop, shoot, tomb…
ɑ ɪ	16. Price	Ripe, write, arrive…
ɔ ɪ	17. Choice	Noise, join, toy…
ɑ ʊ	18. Mouth	Out, house, loud…
ɪ ə	19. Near	Beer, sincere, fear…
ɛ ə	20. Square	Care, fair, pear…
ɑ ː	21. Start	Far, sharp, bark…
ɔ ː	22. North	For, war, short…
ɔ ː	23. Force	Four, wore, sport…
ʊ ə	24. Cure	Poor, tourist, pure…

Table A.2 Standard lexical sets (Wells 1982b)

Appendix 2

	/ʌ/ mud	/ ɑː/ path	/ ɑː/ palm	/iː/ hazy	/ɹ/ bar	/ʊ/ pull	/h/ harm	/ɡ/ sing	/j/ few	[eɪ] gate
Scotland & N.Ireland	+	-	-	-	+	-	+	-	+	-
S.Ireland	+	+	+	+	+	+	+	-	+	-
Northeast	-	-	+	+	-	+	+	-	+	-
Central North	-	-	+	-	-	+	-	-	+	-
Central Lancs.	-	-	+	-	+	+	-	+	+	-
Merseyside	-	-	+	+	-	+	-	+	+	+
Humberside	-	-	+	+	-	+	-	-	+	-
NW. Midlands	-	-	+	-	-	+	-	+	-	+
E. Midlands	-	-	+	-	-	+	-	-	+	+
S. Midlands	+	+	+	+	-	+	-	+	-	+
E. South-west	+	-	-	+	+	+	-	-	+	+
W.South-west	+	-	-	+	+	+	-	-	+	-
South-east	+	+	+	+	-	+	-	-	+	+
East Anglia	+	+	+	+	-	+	+	-	-	+
Wales	+	-	+	+	-	+	-	-	+	-

Table A.3 Main UK regional accent differences expressed as key phonological characteristics (Hughes, Trudgill & Watt, 2005)

Appendix 3

Speechreading Questionnaire

Accent Variation and Lip-reading

Amy Irwin, Institute of Hearing Research, University Park, Nottingham

Personal Information

Name:... ... ...

Age:... ... ...

Place of Birth:... ... ...

1. Please rate your level of hearing ability:

Good	Partially Impaired	Severely Impaired	Deaf
☐	☐	☐	☐

2. Please state the region in which you live (e.g. Nottinghamshire, Derbyshire etc.)

... ...

Accent

3. If you think you speak with an accent please name it here (i.e. Glaswegian etc.)

... ...

4. Would you say your accent was strong enough to be recognised?: Yes ☐ No ☐

5. When conversing with others have you ever found that a persons accent can make it more difficult to understand what they are saying?

Yes, always	Yes, sometimes	Yes, occasionally	No, never
☐	☐	☐	☐

6. Give 3 examples of British accents you find difficult to understand:

1. ... ... 2... ... ... 3 ... ...

7. Please indicate, by ranking in order of importance from 1 to 7 (1 having the most impact), those factors you think would make it more **difficult** for you to understand speech;

Accent type (i.e. where the person comes from)... ... ..

The strength of their accent ... .

Fast rate of speech ... . Slow rate of speech ... .

Mumbling... ... . Dialect (i.e. the use of regional slang) ... .

Low volume of speech (i.e. someone who whispers) ... .

Lip-reading

7. Is lip-reading a skill that you use?

Yes, frequently ☐ Yes, Sometimes ☐ Yes, occasionally ☐ No, never ☐ Go to 15.

8. How many years of lip-reading experience do you have? ... ...

9. Please rate your level of lip-reading ability: Good ☐ Average ☐ Poor ☐

10. When lip-reading are you able to identify the speakers accent?

Yes, always ☐ Yes, sometimes ☐ Yes, occasionally ☐ No, never ☐ Go to 12.

11. Please state the three British accents you have found the most obvious, or easy, to identify through lip-reading: 1... ... ... 2... ... ...

3... ... ...

12. When lip-reading do you find that a person's accent can make it more difficult to understand them?

Yes, always ☐ Yes, sometimes ☐ Yes, occasionally ☐ No, never ☐ Go to 15.

13. Please state the three accents (British or other) you find most difficult to lip-read:

1... ... ... 2... ... ... 3... ... ...

14. In your opinion, what is it about certain accents that makes them more difficult to lip-read? ... ...

....

Your Say

15. Please use the space below to add any further comments you may have regarding accent and comprehension of speech:

... ...

Appendix 4

(I) talkers	(J) talkers	Mean Difference (I-J)	Std. Error	Sig.(a)	95% Confidence Interval for Difference(a) Lower bound	Upper bound
1	2	-.458	.180	1.000	-1.158	.241
	3	-1.167	.311	.068	-2.373	.039
	4	-1.542(*)	.371	.025	-2.981	-.102
	5	-1.750(*)	.439	.039	-3.454	-.046
	6	-2.208(*)	.466	.006	-4.017	-.400
	7	-2.833(*)	.461	.000	-4.622	-1.045
	8	-3.333(*)	.661	.003	-5.899	-.768
	9	-3.958(*)	.774	.002	-6.963	-.954
	10	-4.292(*)	.793	.001	-7.368	-1.216
	11	-5.083(*)	.830	.000	-8.302	-1.865
	12	-5.708(*)	.898	.000	-9.192	-2.225
2	3	-.708	.364	1.000	-2.119	.702
	4	-1.083	.345	.305	-2.423	.257
	5	-1.292	.423	.374	-2.934	.351
	6	-1.750(*)	.422	.026	-3.389	-.111
	7	-2.375(*)	.416	.001	-3.990	-.760
	8	-2.875(*)	.626	.009	-5.306	-.444
	9	-3.500(*)	.747	.007	-6.398	-.602
	10	-3.833(*)	.758	.003	-6.775	-.892
	11	-4.625(*)	.834	.001	-7.860	-1.390
	12	-5.250(*)	.837	.000	-8.497	-2.003
3	4	-.375	.454	1.000	-2.135	1.385
	5	-.583	.466	1.000	-2.391	1.224
	6	-1.042	.472	1.000	-2.873	.790
	7	-1.667	.441	.064	-3.378	.044
	8	-2.167	.642	.172	-4.657	.323
	9	-2.792	.732	.059	-5.633	.049
	10	-3.125(*)	.790	.042	-6.191	-.059
	11	-3.917(*)	.843	.007	-7.186	-.648
	12	-4.542(*)	.913	.003	-8.083	-1.000
4	5	-.208	.504	1.000	-2.162	1.745
	6	-.667	.374	1.000	-2.119	.786
	7	-1.292	.512	1.000	-3.280	.697
	8	-1.792	.584	.357	-4.056	.472
	9	-2.417	.725	.190	-5.228	.395
	10	-2.750(*)	.703	.046	-5.476	-.024
	11	-3.542(*)	.816	.016	-6.709	-.374
	12	-4.167(*)	.800	.002	-7.271	-1.062
5	6	-.458	.625	1.000	-2.885	1.969
	7	-1.083	.551	1.000	-3.223	1.056
	8	-1.583	.634	1.000	-4.043	.876
	9	-2.208	.643	.149	-4.702	.285
	10	-2.542	.740	.149	-5.411	.328
	11	-3.333	.861	.051	-6.674	.008
	12	-3.958(*)	.829	.005	-7.173	-.744
6	7	-.625	.469	1.000	-2.446	1.196

	8	-1.125	.553	1.000	-3.269	1.019
	9	-1.750	.663	.965	-4.322	.822
	10	-2.083	.631	.206	-4.532	.365
	11	-2.875	.817	.122	-6.046	.296
	12	-3.500(*)	.749	.007	-6.408	-.592
7	8	-.500	.532	1.000	-2.563	1.563
	9	-1.125	.718	1.000	-3.911	1.661
	10	-1.458	.678	1.000	-4.090	1.173
	11	-2.250	.725	.332	-5.065	.565
	12	-2.875(*)	.718	.037	-5.661	-.089
8	9	-.625	.492	1.000	-2.534	1.284
	10	-.958	.348	.750	-2.310	.393
	11	-1.750	.539	.236	-3.842	.342
	12	-2.375(*)	.548	.016	-4.500	-.250
9	10	-.333	.537	1.000	-2.418	1.751
	11	-1.125	.726	1.000	-3.940	1.690
	12	-1.750	.519	.173	-3.762	.262
10	11	-.792	.611	1.000	-3.162	1.578
	12	-1.417	.524	.841	-3.452	.618
	12	-.625	.766	1.000	-3.596	2.346
11	1	5.708(*)	.898	.000	2.225	9.192
	2	5.250(*)	.837	.000	2.003	8.497
	3	4.542(*)	.913	.003	1.000	8.083
	4	4.167(*)	.800	.002	1.062	7.271
	5	3.958(*)	.829	.005	.744	7.173
	6	3.500(*)	.749	.007	.592	6.408
	7	2.875(*)	.718	.037	.089	5.661
	8	2.375(*)	.548	.016	.250	4.500
	9	1.750	.519	.173	-.262	3.762
	10	1.417	.524	.841	-.618	3.452

Based on estimated marginal means
* The mean difference is significant at the .05 level.
a Adjustment for multiple comparisons: Bonferroni.

Table A.4.1 Pairwise comparisons for Glaswegian talkers (1-6) and Nottingham talkers (7-12), for Nottingham participants (Experiment 5)

(I) talkers	(J) talkers	Mean Difference (I-J)	Std. Error	Sig.(a)	95% Confidence Interval for Difference(a)	
					Lower bound	Upper bound
1	2	-.529	.259	1.000	-1.603	.544
	3	-1.118	.352	.390	-2.579	.344
	4	-1.176	.431	.977	-2.963	.610
	5	-2.059	.525	.080	-4.236	.118
	6	-2.706(*)	.520	.006	-4.862	-.550
	7	-1.824	.502	.147	-3.904	.257
	8	-2.176	.676	.355	-4.982	.629
	9	-4.059(*)	.909	.026	-7.830	-.287
	10	-1.706	.498	.230	-3.772	.361
	11	-3.000(*)	.707	.041	-5.933	-.067
	12	-3.882(*)	.866	.025	-7.472	-.293
2	3	-.588	.470	1.000	-2.538	1.362
	4	-.647	.641	1.000	-3.307	2.013
	5	-1.529	.607	1.000	-4.047	.988
	6	-2.176	.643	.250	-4.843	.490
	7	-1.294	.580	1.000	-3.700	1.112
	8	-1.647	.742	1.000	-4.726	1.432
	9	-3.529	.940	.114	-7.427	.368
	10	-1.176	.698	1.000	-4.071	1.718
	11	-2.471	.782	.400	-5.713	.772
	12	-3.353	.919	.143	-7.166	.460
3	4	-.059	.433	1.000	-1.854	1.736
	5	-.941	.481	1.000	-2.936	1.054
	6	-1.588(*)	.374	.041	-3.141	-.036
	7	-.706	.574	1.000	-3.085	1.674
	8	-1.059	.699	1.000	-3.958	1.841
	9	-2.941	.941	.431	-6.845	.962
	10	-.588	.478	1.000	-2.570	1.394
	11	-1.882	.685	.946	-4.725	.960
	12	-2.765	.983	.828	-6.843	1.314
4	5	-.882	.492	1.000	-2.922	1.157
	6	-1.529	.486	.411	-3.545	.486
	7	-.647	.691	1.000	-3.513	2.219
	8	-1.000	.691	1.000	-3.867	1.867
	9	-2.882	1.043	.913	-7.207	1.442
	10	-.529	.333	1.000	-1.912	.853
	11	-1.824	.734	1.000	-4.867	1.220
	12	-2.706	.865	.428	-6.293	.882
5	6	-.647	.477	1.000	-2.625	1.331
	7	.235	.667	1.000	-2.532	3.002
	8	-.118	.606	1.000	-2.629	2.394
	9	-2.000	.916	1.000	-5.797	1.797
	10	.353	.594	1.000	-2.109	2.815
	11	-.941	.639	1.000	-3.590	1.708
	12	-1.824	.963	1.000	-5.819	2.172
6	7	.882	.652	1.000	-1.823	3.588
	8	.529	.536	1.000	-1.695	2.754
	9	-1.353	.717	1.000	-4.327	1.621

	10	1.000	.374	1.000	-.550	2.550
	11	-.294	.679	1.000	-3.112	2.523
	12	-1.176	.880	1.000	-4.825	2.472
7	8	-.353	.786	1.000	-3.611	2.905
	9	-2.235	.802	.872	-5.562	1.092
	10	.118	.624	1.000	-2.469	2.704
	11	-1.176	.583	1.000	-3.595	1.242
	12	-2.059	.976	1.000	-6.105	1.988
8	9	-1.882	.762	1.000	-5.041	1.276
	10	.471	.563	1.000	-1.865	2.806
	11	-.824	.620	1.000	-3.394	1.747
	12	-1.706	.706	1.000	-4.633	1.222
9	10	2.353	.836	.820	-1.112	5.818
	11	1.059	.774	1.000	-2.151	4.269
	12	.176	.896	1.000	-3.540	3.893
10	11	-1.294	.617	1.000	-3.853	1.265
	12	-2.176	.773	.820	-5.382	1.029
11	12	-.882	.874	1.000	-4.507	2.742

Based on estimated marginal means
* The mean difference is significant at the .05 level.
a Adjustment for multiple comparisons: Bonferroni.

Table A.4.2 Pairwise comparisons for Glaswegian talkers (1-6) and Nottingham talkers (7-12), for Glaswegian participants (Experiment 5)

Appendix 5

(I) Ntal ker	(J)N talker	Mean Difference (I-J)	Std. Error	Sig.(a)	95% Confidence Interval for Difference(a)	
					Lower bound	Upper bound
1	2	2.083	1.041	.071	-.207	4.374
	3	4.083(*)	.733	.000	2.470	5.697
	4	4.083(*)	.830	.000	2.257	5.910
	5	4.167(*)	.920	.001	2.142	6.191
	6	5.167(*)	1.014	.000	2.935	7.398
2	1	-2.083	1.041	.071	-4.374	.207
	3	2.000(*)	.879	.044	.065	3.935
	4	2.000	1.073	.089	-.362	4.362
	5	2.083(*)	.925	.046	.048	4.119
	6	3.083(*)	1.018	.012	.842	5.325
3	1	-4.083(*)	.733	.000	-5.697	-2.470
	2	-2.000(*)	.879	.044	-3.935	-.065
	4	.000	.389	1.000	-.857	.857
	5	.083	.657	.901	-1.362	1.529
	6	1.083	.657	.127	-.362	2.529
4	1	-4.083(*)	.830	.000	-5.910	-2.257
	2	-2.000	1.073	.089	-4.362	.362
	3	.000	.389	1.000	-.857	.857
	5	.083	.645	.900	-1.336	1.503
	6	1.083	.499	.053	-.016	2.182
5	1	-4.167(*)	.920	.001	-6.191	-2.142
	2	-2.083(*)	.925	.046	-4.119	-.048
	3	-.083	.657	.901	-1.529	1.362
	4	-.083	.645	.900	-1.503	1.336
	6	1.000	.461	.053	-.014	2.014
6	1	-5.167(*)	1.014	.000	-7.398	-2.935
	2	-3.083(*)	1.018	.012	-5.325	-.842
	3	-1.083	.657	.127	-2.529	.362
	4	-1.083	.499	.053	-2.182	.016
	5	-1.000	.461	.053	-2.014	.014

Based on estimated marginal means
* The mean difference is significant at the .05 level.

Table A.5 Pairwise comparisons for the Nottingham talkers, based on mean performance by Group 2 (Experiment 6)

(I) talkerss	(J) talkerss	Mean Difference (I-J)	Std. Error	Sig.(a)	95% Confidence Interval for Difference(a)	
					Lower bound	Upper bound
1	2	.667	1.117	.563	-1.792	3.125
	3	2.000(*)	.769	.025	.308	3.692
	4	2.750(*)	1.188	.041	.136	5.364
	5	-1.917	1.131	.118	-4.406	.573
	6	-2.250(*)	.740	.011	-3.878	-.622
2	1	-.667	1.117	.563	-3.125	1.792
	3	1.333	.732	.096	-.277	2.944
	4	2.083(*)	.933	.047	.030	4.137
	5	-2.583	1.300	.072	-5.444	.277
	6	-2.917(*)	1.164	.029	-5.479	-.354
3	1	-2.000(*)	.769	.025	-3.692	-.308
	2	-1.333	.732	.096	-2.944	.277
	4	.750	.664	.283	-.712	2.212
	5	-3.917(*)	1.240	.009	-6.646	-1.188
	6	-4.250(*)	1.188	.004	-6.864	-1.636
4	1	-2.750(*)	1.188	.041	-5.364	-.136
	2	-2.083(*)	.933	.047	-4.137	-.030
	3	-.750	.664	.283	-2.212	.712
	5	-4.667(*)	1.519	.011	-8.010	-1.323
	6	-5.000(*)	1.523	.007	-8.351	-1.649
5	1	1.917	1.131	.118	-.573	4.406
	2	2.583	1.300	.072	-.277	5.444
	3	3.917(*)	1.240	.009	1.188	6.646
	4	4.667(*)	1.519	.011	1.323	8.010
	6	-.333	1.453	.823	-3.531	2.865
6	1	2.250(*)	.740	.011	.622	3.878
	2	2.917(*)	1.164	.029	.354	5.479
	3	4.250(*)	1.188	.004	1.636	6.864
	4	5.000(*)	1.523	.007	1.649	8.351
	5	.333	1.453	.823	-2.865	3.531

Based on estimated marginal means
* The mean difference is significant at the .05 level.

Table A.5.2: Pairwise comparisons for Nottingham talkers (1 & 2), Glaswegian talkers (3 & 4) and Southampton talkers (5 & 6), based on mean performance by group 3 (Experiment 6)

Appendix 6

(I) talker	(J) talker	Mean Difference (I-J)	Std. Error	Sig.(a)	95% Confidence Interval for Difference(a) Lower bound	Upper bound
1	2	.071	.829	1.000	-3.550	3.693
	3	-2.286	1.045	1.000	-6.854	2.283
	4	-4.643	1.436	.431	-10.918	1.633
	5	-5.286(*)	1.087	.020	-10.035	-.537
	6	-5.929(*)	1.097	.008	-10.722	-1.135
	7	-7.929(*)	1.721	.032	-15.450	-.408
	8	-9.143(*)	2.048	.042	-18.096	-.190
	9	-5.643(*)	1.030	.007	-10.146	-1.140
	10	-4.714	1.126	.070	-9.637	.208
	11	-6.429(*)	1.036	.002	-10.958	-1.900
	12	-8.071(*)	1.656	.020	-15.308	-.835
2	3	-2.357	.746	.496	-5.616	.902
	4	-4.714	1.294	.197	-10.371	.942
	5	-5.357(*)	.716	.000	-8.485	-2.229
	6	-6.000(*)	.825	.000	-9.608	-2.392
	7	-8.000(*)	1.258	.002	-13.498	-2.502
	8	-9.214(*)	1.684	.007	-16.576	-1.852
	9	-5.714(*)	.658	.000	-8.591	-2.837
	10	-4.786(*)	.827	.004	-8.399	-1.173
	11	-6.500(*)	1.093	.003	-11.278	-1.722
	12	-8.143(*)	1.235	.001	-13.542	-2.744
3	4	-2.357	1.189	1.000	-7.553	2.839
	5	-3.000	.896	.345	-6.915	.915
	6	-3.643(*)	.634	.004	-6.415	-.871
	7	-5.643(*)	1.146	.018	-10.654	-.632
	8	-6.857	1.735	.109	-14.439	.725
	9	-3.357(*)	.684	.019	-6.348	-.366
	10	-2.429	1.157	1.000	-7.483	2.626
	11	-4.143	.988	.070	-8.462	.176
	12	-5.786	1.327	.051	-11.584	.013
4	5	-.643	1.203	1.000	-5.899	4.613
	6	-1.286	1.081	1.000	-6.012	3.441
	7	-3.286	1.112	.735	-8.144	1.572
	8	-4.500	1.362	.376	-10.452	1.452
	9	-1.000	1.114	1.000	-5.870	3.870
	10	-.071	1.030	1.000	-4.571	4.429
	11	-1.786	1.285	1.000	-7.400	3.829
	12	-3.429	1.244	1.000	-8.863	2.006
5	6	-.643	1.097	1.000	-5.440	4.154
	7	-2.643	1.151	1.000	-7.675	2.389
	8	-3.857	1.342	.860	-9.722	2.008
	9	-.357	.760	1.000	-3.680	2.966
	10	.571	.837	1.000	-3.086	4.229
	11	-1.143	1.089	1.000	-5.901	3.615
	12	-2.786	.984	.937	-7.088	1.517
6	7	-2.000	1.245	1.000	-7.440	3.440

	8	-3.214	1.770	1.000	-10.951	4.523
	9	.286	.822	1.000	-3.305	3.877
	10	1.214	.973	1.000	-3.039	5.468
	11	-.500	1.073	1.000	-5.189	4.189
	12	-2.143	1.402	1.000	-8.270	3.985
7	8	-1.214	1.054	1.000	-5.823	3.394
	9	2.286	.969	1.000	-1.949	6.520
	10	3.214	1.090	.747	-1.551	7.980
	11	1.500	1.235	1.000	-3.897	6.897
	12	-.143	.994	1.000	-4.486	4.200
8	9	3.500	1.599	1.000	-3.490	10.490
	10	4.429	1.221	.203	-.909	9.766
	11	2.714	1.714	1.000	-4.778	10.207
	12	1.071	.917	1.000	-2.935	5.078
9	10	.929	.917	1.000	-3.078	4.935
	11	-.786	.820	1.000	-4.369	2.798
	12	-2.429	1.261	1.000	-7.940	3.083
10	11	-1.714	1.211	1.000	-7.007	3.578
	12	-3.357	1.151	.794	-8.389	1.675
11	12	-1.643	1.389	1.000	-7.714	4.429

Based on estimated marginal means
* The mean difference is significant at the .05 level.
a Adjustment for multiple comparisons: Bonferroni.

Table A.6: Pairwise comparisons for Glaswegian talkers (1-6) and Nottingham talkers (7-12) (Experiment 8)