Identification of language and accent through visual speech

Amy Irwin, Sharon Thomas

Institute of Hearing Research, University Park,
Nottingham, NG7 2RD
{amy; sharon}@ihr.mrc.ac.uk

Abstract

The production of speech involves an individual's control of their various articulators (lips, tongue, larynx etc.) to produce auditory speech signals [1]. These movements can be utilised in the processing of visual speech and form the basis of speechreading. However, the production of speech by different talkers can be variable; physiology, accent and speech rate can all change the appearance of the visual signal. The focus of this report is an investigation into the effects of language and accent variation on speechreading, an area previously lacking in systematic research.

Results from two experiments indicate, firstly, that the visual differences between French and English, (both accent and language) can be discriminated through visual speech. Secondly, in a comparison of speechreading performance, English sentences produced using a French accent were found to be significantly more difficult to speechread by English observers than those produced in an English accent.

This research indicates the importance of further study into the effects of accent on speechreading.

1. Introduction

Speech perception is bi-modal, in that both auditory and visual signals are produced. Speechreading refers to the ability to perceive speech based on the visual signals alone. For deaf and hearing impaired individuals, speechreading is an important skill that vastly improves communication in a predominantly hearing society. In addition to this, research also shows that visual speech information has an impact upon hearing observers' auditory speech perception. For example, presentation of a talkers face improves perception of speech in noise [2]. Visual information has also been shown to combine with the auditory signal to induce mistaken interpretation of a signal when a participant is presented with incongruent stimuli [3]. For example, when presented with the auditory token /ba/ and the visual token /ga/ it was shown that most individuals will report perceiving /da/ - a token presented in neither modality but representing a blend of information from both. The results of such studies indicate that information from the visual modality can be used to augment the auditory signal when that signal is degraded and that visual speech has an impact on auditory speech perception even when the auditory signal is clear.

Despite this, speechreading is a difficult skill to acquire, with ability being extremely variable between individuals. This difficulty must, in part, arise from the variability inherent in individual talker's and their modes of speech production (including speed, pronunciation, accent, prosody and lexical stress). The purpose of this study was to examine the effects of a talkers language and accent on speechreading, beginning with a study of language and accent discrimination, followed by an investigation into the effect of accent variation on speechreading ability.

1.1 Normalization

Normalization in speechreading involves the observer extracting all the talker-specific characteristics (i.e. skin tone, eye colour) and retaining only those relevant to phonetic material (i.e. lip movements, position of jaw). Past research [4] has investigated the effects of normalization by comparing observers' speechreading abilities when presented with either a single talker or multiple talkers (varied by trial). It was found that the use of multiple talkers negatively affected speechreading performance, as observers presented with the single speaker responded with greater accuracy than those in the multiple speaker condition. These results suggest that visual information relating to the moving face of a specific talker is retained for some time, inflicting processing costs when the observer is faced with multiple sets of talker information.

In terms of accent variation, it could be argued that different accents produce specific differences in the movements of speech 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 memory. When applied to speechreading performance, it could be hypothesised that different accents will be associated with a decrease in accuracy until the observer had become accustomed to the changes in facial motion.

1.2 Accent and Language

Part of the observed differences between languages and accents arises from differing use of lexical stress. Research indicates that stressed syllables are those that are most clearly articulated within a word [5]. That is, they tend to be higher in pitch, greater in intensity and longer in duration when compared

to the other syllables within a word [6]. In terms of visual speech it has been reported that the size and velocity of lip movements vary with lexical stress [7]. These differences are all visible when the talker is producing speech and as such may have an impact on visual speaker intelligibility.

Lexical stress can also be a differentiator between languages and accents, for example, in comparison to English (which is "stress timed", where the rhythm of a talkers speech is based on stressed syllables appearing at a roughly constant rate [8]), French is a "syllable timed" language. That is, the syllables in French appear at regular intervals regardless of stress. As such, the stress of each word in the French language tends to fall on the final syllable [8]. Lexical stress is not the only differentiator between languages, for example, differences can be noted in the location of the place of articulation, i.e. French and English production of 's', the former being dental the latter alveolar [9]. Such differences lend weight to anecdotal evidence which suggests that when producing a foreign language the speaker will generate visual information which appears foreign to a native speaker [10].

The observed differences between languages lead to the experimental questions addressed in this paper – can observers identify the language a speaker is producing based on visual information alone? Leading on from that, are observers then able to identify the accent with which the speaker is producing the language? Further, does variation in accent have an impact on speechreading ability? In order to address these questions two studies were developed, the first of which addressed the question of language and accent discrimination.

2. Experiment 1: Accent Discrimination

2.1 Participants and stimuli

Thirty adult participants took part in the study; all were native speakers of English and had normal vision.

The stimuli used were 50 short sentences from the Bamford-Kowal-Bench [11] sentence set, with one video recording made in each of four conditions: English with an English accent, English with a French accent, French with an English accent and French with a French accent. The talker used for the recordings was a bilingual English male who had spent many years living in France.

2.2 Design

Each session contained 200 visual clips of sentences, rotated through the four stated conditions. The rotation was randomised for each participant. There were 2 participant groups that differed only in the aspect they were asked to distinguish – Group 1 were asked to decide which language the speaker was using, Group 2 were asked to decide the accent the speaker was using (French or English).

2.3 Procedure

Each participant was seated at a table directly in front of the view screen and instructed to watch each clip carefully. They were asked to make a judgement as to whether the speaker was talking in French or English, or whether he was speaking with a French or English accent, depending on their group allocation. They were not required to comprehend the individual words, only to make a general decision on language or accent. They used a key pad to make their response.

2.4 Results

Both groups completed the task at a level significantly greater than chance (chance designated as 50% correct) ($p < .005$). Although performance levels in Group 2 (accent) were lower (mean: 115.4) than in Group 1 (mean: 126.8) the difference between the two groups was not significant (F: 3.373, $p > .05$). However, when the performance of the two groups was split according to the nature of the stimulus (language / accent pairings) then several interactions both between and within the groups were found. Figure 1 below shows the mean performance (number of correct responses) of participants in both groups, according to the stimulus pairings.

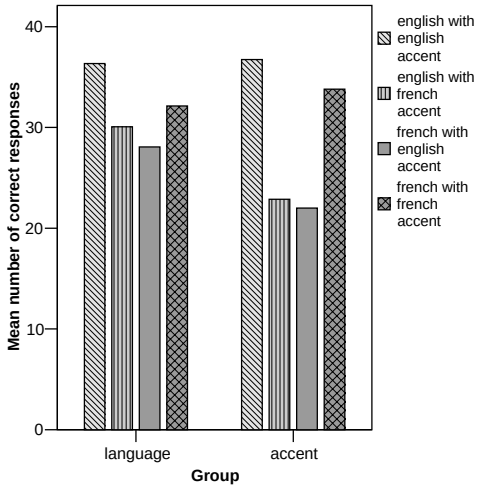

Figure 1: Mean number of correct identifications of language or accent across sentence conditions

Figure 1 shows quite clearly that performance levels in both groups were lower when the stimulus was incongruous (language and accent differed from one another). Indeed, further analysis (one sample t-test) indicated that those in the accent group were operating at chance level ($p < .05$) when the stimulus pairs were incongruous (all other performance levels are above chance). This suggests that identification of both language and accent is easier if the stimulus pair match. Figure 1 also shows that the level of performance when dealing with incongruous stimuli is lower for accent than for language. This suggests that accent is harder to identify than language when the language and accent presented do not match.

The next step in the analysis was to determine if the differences observed were significant. The results indicated that performance was significantly poorer (t: 3.865, sig: .001) in the accent group if they had to identify that the speaker was using a French accent when the language he used was English, in comparison to the language group if they had to identify the language as English when the accent was French. The same was also true when the language was French and the accent used was English – those asked to identify the accent were more likely to respond incorrectly (t: 2.432, sig: .022) than those asked to identify the language. There was no significant difference between the groups when the stimuli were congruent.

Further analysis showed that within Group 1 (language) participants were significantly poorer at identifying language in incongruent stimuli compared to congruent stimuli ($p < .005$). They were also poorer at identifying the language as French, compared to English when the stimuli were congruent ($p < .005$). That result is replicated in Group 2 (accent).

2.5 Discussion

This pattern of errors indicates two effects of language and accent; firstly, when participants are asked to identify the accent, language appears to have a stronger effect than accent in incongruous stimuli. Thus, when language and accent do not match, observers are more likely to respond incorrectly with the language used than respond with the accent. However, when asked to identify the language the speaker is using, participants in Group 1 were poorer at the task when the accent used by the speaker did not match the language. This suggests that accent can also have an effect, by increasing the number of incorrect responses to incongruent stimuli.

In summary, it would appear that individuals are equally capable of identifying accent or language when the stimuli they are presented with are congruent. However, when presented with incongruent stimuli, performance levels in both groups decrease. Thus, both language and accent would seem to have a confusing effect on identification. The effect of language over accent appears to be stronger as the performance of the accent group is significantly worse than the language group when dealing with incongruent stimuli. It could, therefore, be theorised that although accent is difficult to identify when participants are explicitly asked to identify it, it can have an implicit effect on language identification, leading to confusion and larger numbers of incorrect responses when language and accent are presented incongruently.

Having established that French and English accents differ visually, in that accent can have a confusing effect on language identification. The next step was to determine if the differences between the accents would have an effect on speechreading ability.

3. Experiment 2: The effect of accent variation on speechreading ability

3.1 Participants and stimuli

Fourteen adult participants with normal vision and hearing who were native speakers of English were recruited.

A set of 40 BKB sentences, all spoken with an English accent, were presented to assess speechreading ability. A further two sets of 50 sentences were generated, one set spoken with an English accent, one set spoken with a French accent. The speaker and recording conditions were the same as in the previous study.

3.2 Procedure

As before, each participant was seated at a table in front of a computer screen. They were instructed that the speaker would produce 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 into a response box. They were not required to understand the entire sentence. Any word that was typed in was recorded. Each participant completed the speechreading ability test first, before moving on to the experiment proper.

3.3 Results

Each participants score was generated using a loose keyword scoring system where errors in morphology are ignored [11]. Each sentence had 3 or 4 keywords, and a point was awarded for each correctly identified keyword, with closely related examples such as plurals being accepted. A participants score therefore represents the percentage of correctly identified keywords within a sentence set.

Analysis of speechreading accuracy for the experiment proper indicated that the majority of the participants were more accurate when speechreading sentences spoken with an English accent (mean: 26.7% correct) in comparison to those spoken with a French accent (mean: 19.8% correct). A paired samples t-test showed the difference to be significant (t: 4.702, sig: .000).

Following on from that result, the participants were then split into two groups, based on their performance on the speechreading ability test. Mean performance for the test was 27% of keywords correctly identified. 7 participants had scores above the mean and were designated the "good" speechreaders, the remaining 7 participants were designated as "poor" speechreaders. Figure 2 on the following page illustrates the mean percentage of correctly identified keywords by each group.

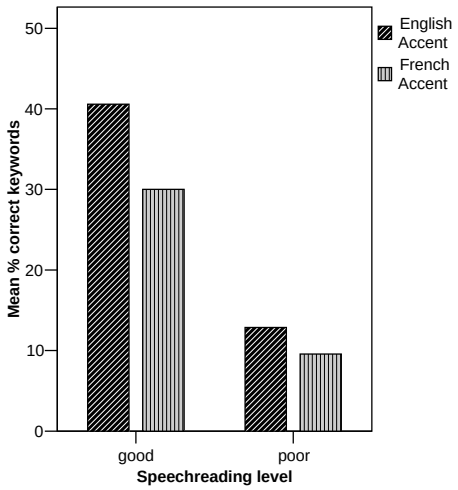

Figure 2: Mean % correctly identified keywords by the "good" and "poor" speechreaders

Both groups were more accurate speechreading when the accent used by the talker was English. Further analysis found that the difference in keyword identification accuracy shown by the good speechreaders was significant (t: 6.632, sig: .001), while the observed difference in performance shown by the poor speechreaders was not (t: 2.137, sig: .077). Finally, an independent samples t-test was conducted to compare the level of difference shown between the groups, this was also found to be significant at the $p < .05$ level (t: 3.354, sig: .006).

3.4 Discussion

The results indicate that speechreading is affected by accent variation. Specifically, the use of a French accent lowers accuracy in identification of keywords by a significant amount. This suggests that the differences in facial movements produced by the less familiar French accent were sufficient to render the visual message less clear and yield lower speechreading accuracy.

Interestingly, the difference in performance is more notable among the better speechreaders in comparison to those whose level of performance was poor. It could be hypothesised that those individuals who are more able to utilise the visual speech signal become accustomed to the general facial movements associated with speech production in an English accent, with the formation of expectations about the appearance of speech on a talkers moving face. As such, the distortion produced when a French accent is used may have more of an impact than if they were unaccustomed to how the message should appear – as in the case of the poor speechreaders. This would confirm anecdotal evidence produced by deaf speechreaders who have reported that they find unfamiliar accents difficult to speechread [10].

4. Conclusion

The results gathered so far indicate strongly that differences in the movement of speech articulators, such as those produced by different accents, do have an effect on speechreading ability.

These results are a preliminary confirmation of the hypothesis that differences in facial movements, related to accent changes, could impair speechreading performance. Further research is required to determine if these effects can be mitigated by the passage of time or through training. As such, the work shown here represents the beginning of a programme designed to both investigate the effects of familiar and unfamiliar accent on speechreading ability and develop training programmes that overcome such effects.

5. References

[1] 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, pp. 1174-1188

[2] 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

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

[4] Yakel, D.A., Rosenblum, L.D. & Fortier, M.A. (2000) Effects of talker variability on speechreading, *Perception and Psychophysics*, 62, 2, pp. 1405-1412

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

[6] O'Shaughnessy, K. (1989) Lexical stress detection in isolated English words, *Speech Communication*, 8, pp. 113-124

[7] Munhall, K.G. & Vakikiotis-Bateson (1998) The moving face during speech communication, *in*, Campbell, R. Dodd, B. & Burnham, D. (eds) (1998) *Hearing By Eye II, Advances in the Psychology of Speechreading and Auditory-Visual Speech*, Psychology Press, London

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

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

[10] Burnham, D. (1998) Language specifity in the development of auditory-visual speech perception, *in*, Campbell, R., Dodd, B. & Burnham, D. (1998) *Hearing By Eye II, Advances in the Psychology of Speechreading and Auditory-Visual Speech*, Psychology Press, London

[11] Bench, J., Kowel, A. & Bamford, J. (1979) The BKB (Bamford-Kowel-Bench) sentence lists for partially hearing children, *British journal of Audiology*, 13, pp. 108-112