Acoustic and visual adaptations in speech produced to counter adverse listening conditions

Valerie Hazan[1], *Jeesun Kim*[2]

[1]Department of Speech Hearing and Phonetic Sciences, University College London, London, UK
2 The MARCS Institute, University of Western Sydney, Sydney, Australia

v.hazan@ucl.ac.uk, J.Kim@uws.edu.au

Abstract

This study investigated whether communication modality affects talkers' speech adaptation to an interlocutor exposed to background noise. It was predicted that adaptations to lip gestures would be greater and acoustic ones reduced when communicating face-to-face. We video recorded 14 Australian-English talkers (Talker A) speaking in a face-to-face or auditory only setting with their interlocutors who were either in quiet or noise. Focusing on keyword productions, acoustic-phonetic adaptations were examined via measures of vowel intensity, pitch, keyword duration, vowel F1/F2 space and VOT, and visual adaptations via measures of vowel inter-lip area. The interlocutor adverse listening conditions lead Talker A to reduce speech rate, increase pitch and expand vowel space. These adaptations were not significantly reduced in the face-to-face setting although there was a trend for a smaller degree of vowel space expansion than in the auditory only setting. Visible lip gestures were more enhanced overall in the face-to-face setting, but also increased in the auditory only setting when countering the effects of noise. This study therefore showed only small effects of communication modality on speech adaptations.

Index Terms: speech adaptation, audiovisual communication, speech in noise.

1. Introduction

Speech communication often occurs in less than ideal conditions, and can be affected by the presence of background noise, other speakers talking, or by speaker-related factors such as hearing loss or poor mastery of the language being spoken. Talkers usually compensate for these different types of interference by adapting their speech production, i.e. by adopting a 'clear speaking style'. Such adaptations involve changes to the acoustic-phonetic characteristics of the speech such as decreases in speech rate [1, 2, 3], increases in the number and duration of pauses [1, 4], increases in overall intensity [5] and in pitch mean and range [1]. There can also be more fine-grained segmental modifications such as vowel hyperarticulation and increased VOT differences in stop voicing contrasts (see [6] for a review). The Hyper-Hypo (H&H) theory of speech production [7] is a useful framework for work investigating speech adaptations of this type as it argues that talkers use the control that they have over their speech production to maximize communication efficiency in different communicative situations; this entails an ongoing tension between a talker's desire to minimize effort and the need to hyperarticulate in order to be understood.

There is evidence that the specific acoustic-phonetic adaptations that talkers make in clear speech are at least partly dependent on the acoustic characteristics of the adverse conditions under which communication is taking place [8]. Such attunements could also be at the level of the weight given to adaptations in different modalities. A recent debate has centered on whether visual cues are also enhanced in clear speech. If there are, are these visual enhancements merely a consequence of increased articulations and therefore linked to acoustic enhancements, or are they due to specific efforts made by speakers to exaggerate visible articulatory gestures when communicating face-to-face? Fitzpatrick, Kim and Davis [9] addressed this question in a study in which they recorded four speakers while they carried out a Sudoku-like task in quiet and in noise. Speech amplitude was significantly lower when interlocutors could see each other while both communicated in noise, and talkers instead increased the saliency of their visual speech production (as measured from lip area) for noisy conditions involving face-to-face communication. This study did not include acoustic measures of vowel space. They concluded that talkers 'actively monitor their environment and adopt appropriate speech production for efficient communication'. In a study involving a single speaker [10], Garnier et al. analysed the Lombard speech produced when the talker carried out an interactive task in noise in audio or face-to-face conditions or when he did the task alone. In addition to acoustic signals, articulatory lip and tongue movements were also recorded using an electromagnetic articulograph. They investigated whether the visible articulations (e.g. lip spreading, protrusion) were more greatly amplified in face-to-face communication than less visible articulations such as tongue movements. As in [9], Garnier et al. found that acoustic adaptations were greater in AO than in AV conditions, and were also greater in tasks involving an interlocutor than when the task was carried out alone, but, contrary to [9], visible articulatory gestures were not more enhanced in the AV than in the AO condition. Their data therefore countered Fitzpatrick et al.'s view that talkers specifically enhanced their visual gestures in face to face communication.

Both studies involved small numbers of talkers, so the aim of this study was to investigate acoustic-phonetic and visual measures in similar communicative conditions using a larger corpus of 14 talkers. There were some important differences in our study. To focus on the changes that talkers specifically makes for the benefit of their interlocutor rather than as a direct consequence of producing speech in background noise (Lombard effect), in our study only one of the talkers in the task was affected by noise and we investigated the acoustic-phonetic and visual adaptations that the unaffected talker made to maintain effective communication (as in the auditory only study on clear speech reported in [11]). Previous studies with both adults and children have shown that these conditions elicit reliable acoustic-phonetic adaptations in the unaffected talker [11, 12]. Using a game-like interactive task carried out in different conditions, we examined the effect of modality (audio alone, audiovisual) and transmission condition affecting talker B (‘no barrier’, background multibabble noise) on speech produced with communicative intent.

Our key research questions were as follows: what acoustic and visual enhancements occur to counter the effects of adverse listening conditions, and ware acoustic enhancements affected by the presence of visual cues in face-to-face conditions. If face-to-face communication generally makes the task easier, we may expect to see a general reduction in acoustic enhancements in the face-to-face noisy conditions. Furthermore, our prediction is that global (e.g., pitch and intensity) and segmental (e.g., stop voicing contrasts) measures that are not visibly marked will show less effect of modality than segmental measures of vowel production for which there are visual cues.

2. Method

2.1. Participants

Fourteen adults (10 women, 4 men; age range: 19-35 years; mean: 26 years) participated in the study. They all spoke Australian English as their primary language, and around half spoke additional languages.

2.2. Materials

Our aim was to compare the acoustic-phonetic and visual characteristics of a subset of phonetic categories produced in different conditions. A task was constructed to naturally elicit a series of keywords within an interactive game. The keywords ‘red’, ‘black’, ‘blue’, ‘green’ were included to get measures for the vowels /ɛ/, /æ/, /u/, /i/ and the letters B, P, V, D to investigate the production of stop voicing (B vs P), and consonant place of articulation (B vs V vs D). In this paper, only the vowel and stop consonant voicing data are presented.

A ‘SAME/TRAP’ grid task (see Figure 1) was designed to be carried out between two speakers. It was a simpler task than the typically-used Map Task [13] or Diapix task [14, 15] so that it could be carried out in face-to-face conditions with talkers only needing to briefly glance at the picture. The task involves two participants having to find identical cells in two pictures as well as ‘traps’. Each of the two grids consists of 16 cells. Each cell contains a coloured picture and a letter: the letter corresponds to the first letter of the object depicted in the cell and is in the same colour as the picture; participants report each cell as the letter/colour combination (e.g. ‘cell 1 has a green V’). Four out of the sixteen cells are identical in both grids (same letter, colour and object): these are the SAME cells. Four other cells share the same letter and colour but have a different object: these are the TRAP cells. The task of the players is to find all the SAME and TRAP cells and indicate these on the empty grid given to talker A. Participants were told that they must not say the names of the objects directly but must paraphrase to work out whether they have got the same or different objects (‘*Is your ‘green V something you drive?*’). Nine sets of grids were designed so that the task could be repeated in different communicative conditions.

2.3. Procedure

The recording space consisted of two adjacent sound-treated rooms connected by a two-way window and a control room where all the recording equipment was located. Each talker was seated in one of the rooms and an opaque curtain could be pulled across the window in Room B for the Audio alone test conditions; this obscured the faces of the participants but not their upper body. To capture the speaker’s mouth motion, participants wore a lightweight purpose-built camera rig (see Figure 2). This rig consisted of a miniature colour camera that was mounted on the apogee of a lightweight rigid supporting arm attached to an adjustable head-band; also mounted on the supporting arm was a strip of LEDs that were used to provide an even source of illumination (see Figure 2, a). The camera was positioned so that it was directed at the talkers’ lips and mouth (see Figure 2, b) and the arm of the rig was adjusted such that it did not obscure the talkers’ face or restrict movement. This camera setup allowed talkers to move their head freely while a fixed region was maintained for the video capture. Prior to each recording session participants wore the camera rig to familiarise themselves with the setup. Ear-mounted headphones were also worn. Participants faced each other at a slight angle to avoid reflections. Each participant was also lit from below by a spotlight. The grid used in the task was attached on the side of the window close to eye-level.

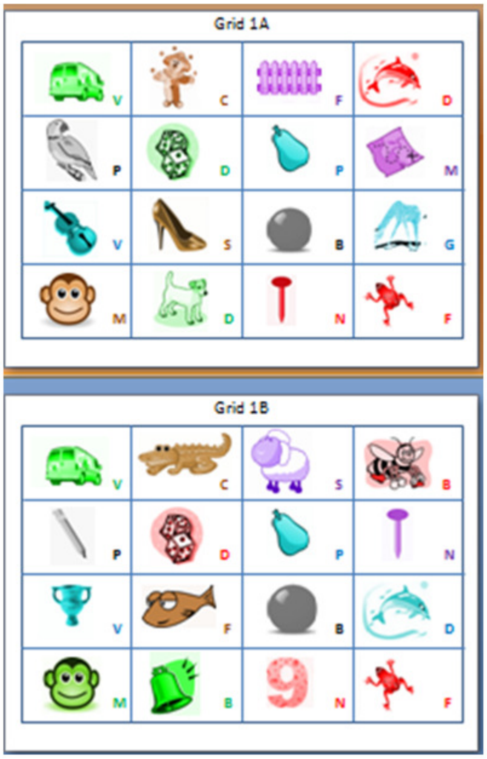
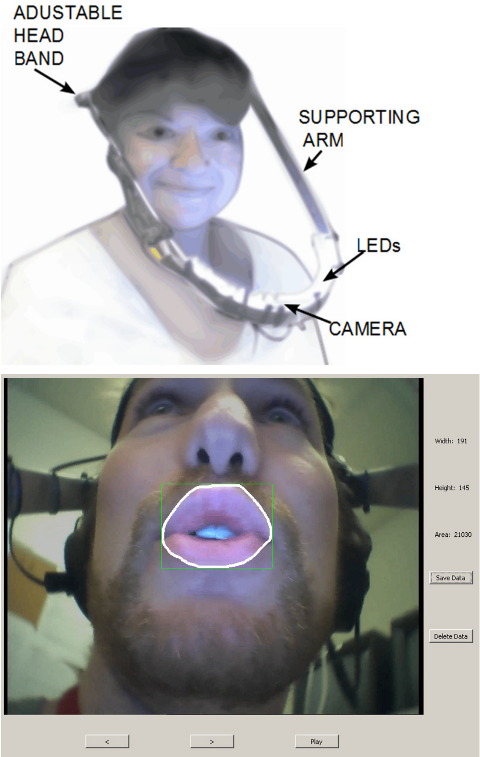


Figure 1: *Example of grids used by talkers A and B in the ‘SAME/TRAP’ task*

The grid task was carried out nine times by a talker pair to evaluate the effect of different factors on speech production for talker A. Talker A was told to take the lead in these interactions. The factors that were controlled in the experiment were: test modality (‘audio only’ AO, ‘audiovisual’ AV), transmission condition affecting talker B (‘no barrier’ NB, ‘vocoder’ VOC, ‘babble’ BAB) and communicative intent (‘with interlocutor’, ‘task carried out alone’). In the AO condition, the two talkers could not see each other while in the AV condition, they carried out the task face to face. In the NB condition, they could hear each other normally, in the VOC condition (not reported here), talker B heard talker A’s voice via a three-channel noise-excited vocoder (as in [10]); in the BAB condition, talker B heard talker A's voice with multitalker babble noise in the background at a fixed level that was set during a pilot phase to elicit comprehension difficulties for talker B. In the communicative conditions, talker A carried out the task with talker B while in the 'alone' condition, talker A had to compare the two pictures aloud and describe the differences.

In summary, the full set of conditions for talker A were as follows: AV block (AV NB, AV VOC, AV BAB), A block (A NB, A VOC, A BAB), Alone block (Alone NB, Alone VOC, Alone BAB). The following factors were counterbalanced across participant pairs: Test modality order: (Order A: AO block, AV block, Alone block; Order B: AV block, AO block, Alone block) and transmission condition order (Order A: NB, VOC, BAB; Order B: NB, BAB, VOC).

In this paper, we focus on the comparison between the two interactive modality conditions (AO, AV) and between two transmission conditions (NB, BAB), so the other conditions are not reported. Data from the following conditions are therefore included in the various acoustic and visual analyses: AO NB, AO BAB, AV NB, AV BAB.



Width: 191

Height: 145

Area: 21030

Figure 2: *Equipment used for recordings: (a) a depiction of the camera rig that was used to capture the mouth region of the participants. (b) software used for tracking outer lip aperture and width.*

2.4. Data processing

Acoustic signal: Three tiers of manual annotations were carried out on the audio files for talker A using Praat software [16]. All annotations were carried out by the first author, a trained phonetician. In the utterance-level tier of the textgrid, all regions containing speech were tagged as SP (thus excluding pauses, noises, laughter, etc), and these were used for the calculation of global measures such as long-term average spectrum. In the keyword-level tier, the beginning and end of all keywords was marked. In the phonetic-level tier, segment boundaries were marked for the vowels and consonants under investigation. The vowel markers included the initial formant transition region. As the keywords were naturally elicited within a spontaneous speech task, the number of times each vowel was produced varied across conditions and participants but on average, talkers produced around 5 repetitions of each keyword within a condition.

Video signal: The video regions containing the vowels to be measured were tracked using the time stamps from the textgrid annotations. Outer-lip width, height and area were calculated for each talker by measuring each frame of the vowel regions in the video files. The analysis was conducted using in-house software in which an adjustable lasso was created and positioned to fit the lip contour (see Figure 2, b); once fitted the outer-lip data was output to a spreadsheet.

2.5. Data analysis

Pitch range and mean: A Praat script was used to calculate measures of median fundamental frequency (F0) and F0 interquartile range in semitones relative to 1 Hz from the complete speech recording per talker per condition.

Keyword duration: A Praat script was used to calculate the duration of each of the words tagged in the keyword tier. Mean keyword duration was then calculated per talker per condition

Vowel intensity: After the recordings had been normalized for peak intensity, the same Praat script as in [9] was used to calculate vowel intensity for the vowels tagged in the keywords. A mean intensity value was calculated per vowel, and then an intensity value averaged over the four vowels was calculated per talker per condition.

Stop consonant voicing: Voice onset time (VOT), the time between burst release and onset of voicing was calculated using a Praat script from annotations in the phonetic-level tier for 'P' and 'B' keywords produced in the task. A median value was calculated per talker per condition.

Vowel formant measurements: First, vowel formant estimates were obtained using a Praat script from the vowel segments annotated on the phonetic tier (i.e. vowels from keywords 'black', 'green', 'blue', with the formant values estimated at vowel mid-point). Values were checked manually for gross errors such as formant skips; if a value seems out of range, the file was examined in Praat and accurate formant values estimated manually. Formant values were converted to ERB, an auditory scale, using the 'vowels' package in R [17] and mean formant values were calculated per talker per condition per vowel. A measure of vowel area was calculated as described in Neel [18], using Heron's formula for the area calculation of the vowel triangle. The triangle consisted of the Euclidean distances in ERB from (a) /i/ to /æ/, (b) /æ/ to /u/, and (c) /u/ to /i/. First the semi perimeter was calculated using the formula s = (a + b + c)/2; then, the triangle area was calculated by taking the square root of s*(s–a)*(s–b)*(s–c).

Visual measure of lip area: In order to minimize the inclusion of incorrect lip width/height estimates, individual values were checked to remove tokens where width, height or both together with inter-lip area were outside of 1.5 SD of the mean for tokens (per speaker, per vowel, per condition). As in [9], the inter-lip area values were then normalised against the maximum possible area for each vowel for each talker. That is, the 'inter-lip area proportion' measure was referenced to the token produced in any condition by that talker that had the maximum area. A mean inter-lip area averaged over the three vowels /æ/, /i/, /ɛ/ was calculated; these are three vowels where inter-lip area would be expected to increase if they were hyperarticulated.

3. Results

3.1. How were global acoustic measures affected by modality and transmission condition?

For all measures, repeated-measures ANOVAs were carried out with within-subject effects of modality (AO, AV) and transmission condition (NB, BAB). Talkers increased their median pitch in the BAB conditions relative to the NB conditions [F(1,13)=26.91; p<0.001]; the modality by condition just failed to reach significance (p=0.066) but there was a trend for median pitch to increase more in BAB relative to NB conditions in the AO mode (89.9 to 91.2 semitones) than in the AV mode (90.0 to 90.8 semitones). The main effect of modality (AO, AV) was not significant. Mean keyword duration was also longer in BAB (388 ms) than NB (350 ms) conditions [F(1,13)=12.71; p<0.005]. This suggests that talkers reduced their speech rate when their interlocutor was hearing them in noise; however, keyword duration was unaffected by modality. Pitch range and vowel intensity did not vary significantly across either condition or modality.

In summary, the presence of noise affecting talker B did lead talker A to make some global adaptations to her or his speech: on average, talkers increased their pitch and reduced their speech rate in these conditions. However, the availability of visual cues in the face-to-face condition had relatively little effect on these global measures.

Table 1. *Measures of F0 median, F0 interquartile range (both in semitones rel. to 1 Hz), mean keyword duration (in msec) and mean vowel amplitude (in dB)*

measure	AO	AV NB	AO	AV
	NB		BAB	BAB
F0	89.9	90.0	91.2	90.9
median	(4.0)	(4.6)	(3.9)	(4.1)
F0 range	3.0	3.4	3.3	3.2
	(0.8)	(1.1)	(0.9)	(0.9)
keyword	339.5	360.9	387.4	388.5
duration	(60)	(61)	(61)	(72)
Mean	67.4	70.0	70.5	69.3
amp	(6.4)	(3.2)	(3.1)	(3.8)

3.2. How were segmental acoustic measures affected by modality and transmission condition?

Next, we examined whether vowel and consonant segmental contrasts within the keywords varied as a function of modality or condition. Our predictions were that acoustic contrasts would be enhanced to counter the adverse listening condition, but that less acoustic enhancement may occur for vowels in the AV condition due to the availability of visual cues for these vowel contrasts. As stop voicing contrasts are not marked visually, we did not expect modality to affect their production.

a. (/b/-/p/ VOT contrast): Statistics were run on the difference between the median VOT values for /b/ and /p/. There was a significant effect of condition [F(1,12)=11.91; p=0.005], with a greater VOT contrast obtained in BAB (79.9 ms) than NB conditions (70.4 ms), but no effect of modality. When the statistical evaluation was carried out on proportional values (/p/ as a proportion of /b/), the effect of condition was only marginally significant (p=0.058), so the increase in VOT contrast in BAB is at least partly linked to the longer keyword durations found in these conditions. These findings confirm expectations that the voicing contrast would not affected by the presence or absence of visual cues.

b. Vowel contrasts: Next, we examined the area measures obtained for the /i/-/æ/-/u/ F1/F2 vowel space (Figure 3), to investigate whether talkers hyperarticulated their vowels when attempting to clarify their speech for the benefit of their interlocutor, as in [11]. Repeated-measures ANOVAs showed a significant effect of condition [F(1,13)=11.53; p=0.005]: the vowel space was expanded in BAB relative to NB conditions. The effect of modality was not significant. We had predicted that less acoustic enhancement would occur in the AV condition due to the availability of visual cues. As can be seen in Figure 3, there was a trend for a greater difference between NB and BAB conditions in AO (4.97 vs 6.67 ERB^2) than AV (5.72 vs 6.17 ERB$^{2)}$, but the modality by condition interaction failed to reach significance (p=0.078).

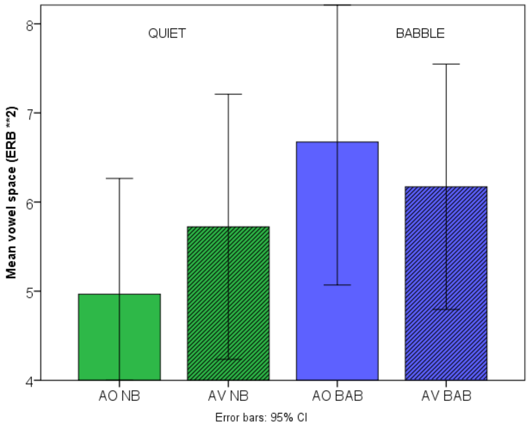


Figure 3: *Measures of vowel area as defined by the Euclidean distance between the vowel /i/, /ae/ and /u/ extracted from the keywords 'green', 'black' and 'blue'.*

3.3. How were visual measures of inter-lip area affected by modality and transmission condition?

Visual measures were then examined to see whether visible lip gestures were more enhanced in face-to-face conditions, which would give support to the view that adaptations were attuned to the specific modality in which the interaction was taking place, given that the acoustic vowel space was not significantly expanded in AV relative to AO.

Repeated-measures ANOVAs were carried out on the mean proportion of inter-lip area averaged over the vowels /æ/, /i/, /ɛ/. The vowel /u/ was not used for this calculation as a visual enhancement of this vowel would be linked to a reduction rather than increase in inter-lip area. The effects of modality [F(1,12)=4.82, p<0.05] and condition [F(1,12)=12.28, p<0.005] were both significant, with no significant interactions. For modality, inter-lip area was larger in AV (0.73) than in AO (0.69). For condition, inter-lip area was larger in BAB (0.76) than NB (0.66).

Table 2. *Mean inter-lip area per condition, calculated per talker as a proportion relative to the maximal area measures for that talker per vowel.*

	AO NB	AV NB	AO BAB	AV BAB
/ae/	.62 (.15)	.65 (.19)	.69 (.17)	.74 (.12)
/i/	.66 (.13)	.72 (.14)	.77 (.07)	.79 (.09)
/e/	.65 (.13)	.73 (.12)	.79 (.11)	.78 (.11)
all	**.63 (.12)**	**.69 (.13)**	**.74 (.09)**	**.77 (.06)**

3.4. Is there evidence of individual strategies in speaker adaptations for vowels?

As individual talkers vary in the strategies they use to clarify their speech [e.g., 19], it is worth examining the individual data for vowel measures in more detail to see whether visual enhancement of lip gestures was a strategy that was only used by some listeners. In Figure 4, the scatterplot shows the mean inter-lip area per mode for each individual talker. Points close to the diagonal (or below it) represent talkers who do not enhance their visual gestures in the AV mode; it therefore appears that only a subset of talkers was using this strategy. The scatterplot showing change in acoustic vowel space for individual talkers (Figure 5) show a stronger correlation across modalities.

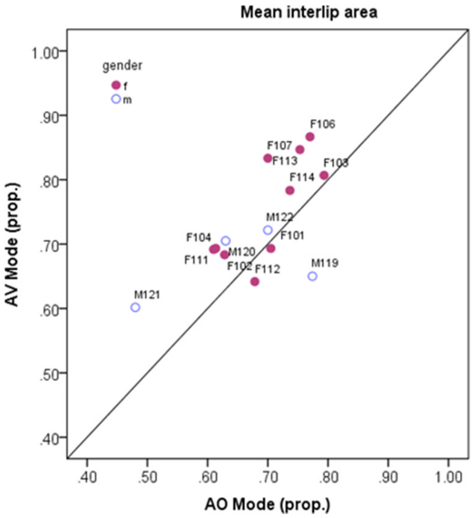


Figure 4: *Scatterplot showing the mean inter-lip area for the AO conditions against that for the AV conditions for individual talkers. Points above the diagonal represent talkers who increase their visible lip gestures in the AV mode.*

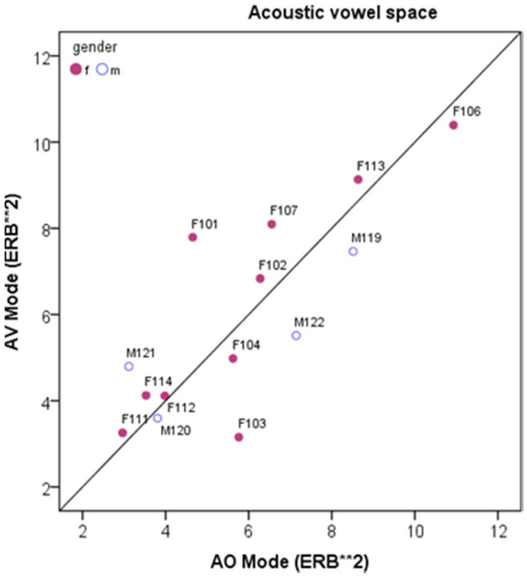


Figure 5: *Scatterplot showing the mean acoustic vowel space area averaged over the two AO conditions against that averaged over the two AV conditions for individual talkers. Points above the diagonal represent talkers who expand their acoustic vowel space in AV.*

4. Discussion

This study investigated some adaptations that talkers make to their speech in order to maximize communication effectiveness in different communicative situations. More specifically, it investigated the degree to which acoustic and visual adaptations varied as a function of the communication modality (whether auditory alone or face to face) when interacting in good listening conditions or with an interlocutor affected by a noisy background. Our predictions were that acoustic adaptations would be reduced in face-to-face situations, due to the presence of additional visual cues, and that visual gestures would be enhanced in face-to-face situations and not simply be a by-product of acoustic enhancements.

Our data did not entirely conform to these expectations. As in our previous study (in A mode) using a different problem-solving task [11], the ‘unimpaired’ talkers slowed down their speech, increased their mean pitch and expanded their vowel space when interacting with an interlocutor hearing them in noise. This again supports the view that speech adaptations are guided by the needs of the interlocutor with the aim of maintaining effective communication, as suggested by Lindblom’s Hyper-Hypo model of speech production. However, here, there was no significant increase in intensity or pitch range. Given the number of talkers tested here, the lack of an effect could be linked to differences in the strategies used by individual talkers to clarify their speech [19] or to the level of difficulty of the task, as discussed below.

A key objective of this study was to investigate whether the availability of visual cues in the face-to-face condition affected the types of adaptations made by talkers. As expected, face-to-face communication had relatively little effect on global acoustic measures which are not much marked visibly (pitch characteristics, intensity, duration). In noise, these global acoustic adaptations were not significantly reduced when visual cues were present. The presence of visual cues also had no impact on the realization of a segmental contrast that is not marked visually (stop voicing contrast). However, in the vowel measures, which were the focus of our investigation, there was a trend for the acoustic vowel space to show a greater degree of adaptation across the NB and BAB conditions in the AO than AV conditions, but the main effect of modality was not significant.

Given that vowel formant space did not vary significantly across the auditory alone and face-to-face conditions, it was of particular interest to see whether the visual gestures for the key vowels (measured as normalized inter-lip area) would be enhanced in face-to-face communication. This was indeed the case as the main effect of modality was significant: even in good listening conditions, talkers visibly articulated their vowels to a greater degree in face-to-face than in auditory-alone communication. However, inter-lip area also increased in the auditory mode in the BAB relative to NB condition, suggesting that this increase in visual clarity is at least partly a consequence of hyperarticulation aimed at expanding the acoustic vowel space.

It is worth noting that the relatively small effects obtained may be linked to a number of factors. First, given that there are individual differences in the strategies that individual talkers used to clarify their speech [19], a lack of a significant effect when testing a relatively small population of talkers may be due to only a subset of talkers using this strategy as suggested by our examination of the individual data. Second, the task that was used was easier than problem-solving tasks that have been used in previous studies of this kind, as the colour keywords were not very confusable; as a result, adding face-to-face communication in the noisy conditions may have had less impact than would have been the case if the task had been more difficult.

5. Conclusions

To achieve successful communication in a variety of environments, talkers need to continuously adapt acoustic-phonetic and linguistic aspects of their speech production [9, 11]. Our data suggest that adaptations are, to a degree, suited to the modalities used in the communication, and that hyperarticulation of lip movements are an additional strategy for increasing the salience between vowels; however, in our study at least, these effects of communication modality on speech adaptations were small.

6. Acknowledgements

This work was supported by an IRIS grant (20211.71719) from University of Western Sydney. We thank Hanna Elkofairi, Michael Fitzpatrick, for their contribution to the collection of the data and Leo Chong, Phillip Hancock for their contribution to the analysis of the visual data. We are very grateful to Chris Davis for his advice at all stages of this study.

7. References

[1] Picheny, M. A., Durlach, N. I., and Braida, L. D., “Speaking clearly for the hard of hearing. II. Acoustic characteristics of clear and conversational speech”, J. Speech Hear. Res., 29: 434–446, 1986.

[2] Smiljanic, R. and Bradlow, A. R., “Production and perception of clear speech in Croatian and English”, J. Acoust. Soc. Am., 118 (3): 1677-1688, 2005.

[3] Uchanski, R. M., Choi, S., Braida, L. D., Reed, C. M., and Durlach, N. I., “Speaking clearly for the hard of hearing. IV. Further studies of the role of speaking rate”, J. Speech Hear. Res., 39: 494–509, 1996.

[4] Liu, S., and Zeng, F.-G., “Temporal properties in clear speech perception”, J. Acoust. Soc. Am., 120: 424–432, 2006.

[5] Krause, J. C., and Braida, L. D., “Acoustic properties of naturally produced clear speech at normal speaking rates”, J. Acoust. Soc. Am., 115:362–378, 2004.

[6] Smiljanic, R. and Bradlow, A., “Speaking and hearing clearly: Talker and listener factors in speaking style changes”, Language and Linguistics Compass, 3 (1): 236-264, 2009.

[7] Lindblom, B., “Explaining phonetic variation: a sketch of the H&H theory”. in W. J. Hardcastle and A. Marchal [eds] Speech Production and Speech Modelling, The Netherlands: Kluwer Academic, pp. 403-439, 1990.

[8] Cooke, M., and Lu, Y., “Spectral and temporal changes to speech produced in the presence of energetic and informational maskers,” J. Acoust. Soc. Am., 128:2059–2069, 2010.

[9] Fitzpatrick, M., Kim, J. and Davis, “The effect of seeing the interlocutor on auditory and visual speech production in noise”, International Conference on Audio-Visual Speech Processing, Volterra, Italy, 2011.

[10] Garnier, M., Ménard, L., Richard, G., “Effect of being seen on the production of visible speech cues. A pilot study on Lombard speech”, Proceedings of Interspeech, Portland, 2012.

[11] Hazan, V. and Baker, R., “Acoustic-phonetic characteristics of speech produced with communicative intent to counter adverse listening conditions”, J. Acoust. Soc. Am. 130: 2139-2152, 2011.

[12] Pettinato, M. and Hazan V., “The development of clear speech strategies in 9-14 year olds”, Proceedings of Meetings on Acoustics, joint ICA/ASA meeting, Montreal 2-7 June, in press

[13] Anderson, A., Bader, M., Bard, E., Boyle, E., Doherty, G. M., Garrod, S., Isard, S., Kowtko, J., McAllister, J., Miller, J., Sotillo, C., Thompson, H. S. & Weinert, R., “The HCRC Map Task Corpus”, Language and Speech, 34: 351-366, 1991.

[14] Van Engen, K. J., Baese-Berk, M., Baker, R. E., Choi, A., Kim, M. and Bradlow, A. R., “The Wildcat Corpus of Native- and Foreign-Accented English: Communicative efficiency across conversational dyads with varying language alignment profiles”, Lang. Speech, 53: 510-540, 2010.

[15] Baker, R., and Hazan, V., “DiapixUK: a task for the elicitation of spontaneous speech dialogs”, Behav. Res. Meth., 43: 761-770, 2011.

[16] Boersma, P. & Weenink, D., “Praat: doing phonetics by computer” [Computer program]. Version 5.3.07, http://www.praat.org/. (last retrieved 5 March 2012.)

[17] Kendall, T. and Thomas, E.R. “Vowels: Vowel Manipulation, Normalization, and Plotting”, R package [CRAN], version 1.2., 2012.

[18] Neel, A.T., “Vowel space characteristics and vowel identification accuracy”, Journal of Speech-Language-Hearing Research, 51: 574-585, 2008.

[19] Ferguson, S.H., & Kewley-Port, D., “Talker differences in clear and conversational speech: Acoustic characteristics of vowels,” Journal of Speech, Language, and Hearing Research, 50: 1241-1255, 2007.