Proceedings of the Institute of Acoustics

THE INFLUENCE OF SPEAKER NORMALISATION AND TRANSMISSION CHANNEL COMPENSATION ON VOWEL IDENTIFICATION IN NATURAL SPEECH

AJ Watkins & SK Boegli

Department of Psychology, Reading University, Reading RG62AL

1. INTRODUCTION

Information beyond the syllable appears not to be essential for vowel identification as listeners are good at identifying isolated vowels [1]. However other experiments suggest that this 'syllable extrinsic' [2] information does make some contribution. The influence of a carrier phrase on a test sound has been widely investigated. The idea is that with different talkers different reference frames are utilised in vowel identification [3]. When shifts in the first formant of a carrier phrase signal a different speaker, the perceptual identity of a subsequent vowel is changed [4, 5]. However, equating the long term average spectra of formant-shifted carriers eliminates these perceptual effects, thus a compensation for averaged spectral characteristics is sufficient to account for these types of 'speaker normalisation' result [5]. Long term spectrum compensation is normally associated with adjustments for the spectral envelope distortion when a sound is transmitted from a source to a listener [6, 7].

Effects of 'syllable extrinsic' information on vowel quality, have been attributed to the use of synthetic speech. However, some experiments have used natural voices, finding that the identity of a male speaker's vowel can be changed by embedding it in a child's carrier [8, 9].

Here we use natural voices and attempt to change the identity of a male and a female speaker's vowels by embedding them in each others sentences (experiment 1). Experiment 2 investigates whether equalising the long term average spectra of these carrier sentences eliminates perceptual changes to the vowels, this is an attempt to replicate the synthetic speech finding with natural speech. The third experiment uses reversed carriers to ask whether phonetic or auditory mechanisms are at work: a reversed speech carrier does not conform to the constraints of natural utterances and might reduce any effects of a phonetic mechanism.

2. GENERAL METHOD

A male and female were recorded rapidly saying various /pVt/ test words in the sentence frame "Please say _____ for me". The male imitated the female in speaking rate and stress pattern. Two types of recordings were made for the male's carrier sentence, one spoken in his normal pitch and one in which he mimicked the female's pitch. Mixed speaker sentences were created by embedding the male test words in the female speaker's carrier sentence, and by embedding female test words in both the original male and the male imitating female carrier sentences. In informal listening only mixed speaker sentences where the carrier was the male imitating the female were heard as wholly originating from a single speaker.

Recordings were made in an IAC 1201 booth using an Sennheiser MKH 40 microphone. These sounds were amplified (Revox A77), low passed filtered at 9 kHz with a 48-dB per octave cutoff slope (Kemo VBF8), digitised with a 16 bit resolution at a sampling frequency of 20 kHz (Data Translation DT2823) and stored with the ILS program RDA running on a Victor PC286 computer. Digital waveforms of the carrier sentences and test words were transferred to a Sun Sparcstation computer for processing. The vowels from the test words were analyzed to find the first and second formants. FFT's were made of hanning windowed segments of each vowel, and a lattice linear prediction (LP) analysis was used to approximate the vocal tract response. For the male speaker the LP analysis was obtained from a 15-ms frame and for the female

speaker it was obtained from a 4.5-ms frame - this kept 1-2 pitch pulses within each speaker's frame. A LP analysis was run every 5-ms throughout the vowels to obtain time averaged values of F1 and F2. These values were used to plot the speakers' vowel spaces. The test words were created by adding a /p_t/ from each speaker's /pɪt/ (a test word not selected for the main experiments) to each of the chosen vowels from that speaker. The test words created were: /pɒt/, /pʌt/, /pæt/, /pɛt/, /pɑt/, and /pɜt/ ("pot", "putt", "pat", "pet", "part", and "pert"). The sound pressure level was equated across the carriers and across the test words.

Sounds were delivered to subjects online under the control of the PC286 computer. Analog signals were created from the digital waveforms with a 16-bit resolution at a conversion rate of 20 kHz (Data translation DT2823) using the ILS program LDA. These signals were low-pass filtered at 9 kHz with a 48-dB per octave cut-off slope (Kemo VBF8) and presented monaurally to subjects with Sennheiser HD480 headphones in an IAC 1201 booth. After presenting a trial, the computer waited for the subject to press a response button before recording the response and presenting the next trial. Visual prompts to listen or respond were conveyed on the computer's screen. A minimum inter-trial interval of 4 seconds was enforced. The duration of each experiment was around 45 minutes. Each subject was given a different random trial order and a different group of 12 subjects were used in each experiment.

3. EXPERIMENT 1

This experiment asks whether vowels from one speaker, which overlap with different vowels in the other speaker's vowel space, are correspondingly misinterpreted when in a mixed speaker sentence. From the overlaps in the vowel spaces, the predictions are as follows. When the male speaker's test word is embedded in the female carrier sentence /pɛt/ will be perceived as /pʌt/, /pæt/ will be perceived as /pɑt/, and both /pʌt/ and /pɑt/ will be perceived as /pɒt/. When the female speaker's test word is embedded in the male and male-imitate carrier sentences /pʌt/ will be perceived as /pɛt/, /pɑt/ will be perceived as /pæt/, and /pɒt/ will be perceived as /pʌt/ or /pɑt/.

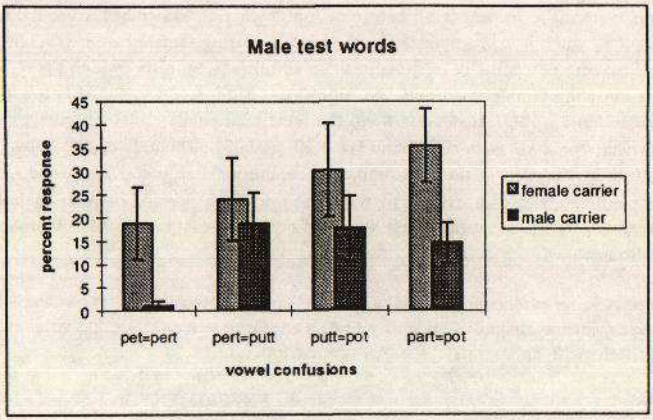
The female test words were presented within the female, the male and the male imitating female carrier sentence, while the male test words were presented within the male and the female carrier sentence. Each of these carrier and test word combinations was presented 8 times giving 6 vowels x 8 repetitions x 5 combinations = 240 trials for each subject. Each subject identified the test word by pressing one of 6 labelled buttons.

Results:

a) Effects of the male and female carriers on errors in the identification of the male test words.

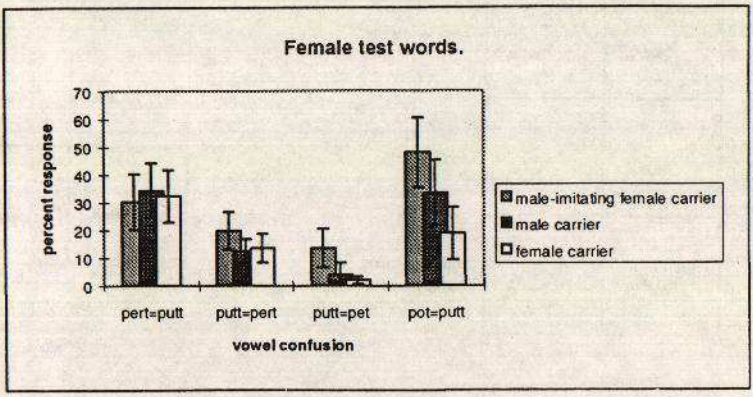
An analysis of variance was carried out on errors made in identification. A main effect of carrier sentence was found [$F(1, 11) = 22.94$, $p<0.001$], indicating that fewer errors were made when the test words were presented within their original male carrier sentence. There was also a main effect of vowel [$F(5, 55) = 7.01$, $p<0.0001$], indicating that the degree of error varied among the test words. An interaction between carrier sentence and vowel [$F(5, 55) = 3.48$, $p<0.01$] indicates that the male and female carriers gave different patterns of error across the test words. Comparisons were carried out in order to investigate the nature of the interaction. Some of these were consistent with the predictions. Significantly more errors were made in the identification of /pɑt/ [$t=3.62$, $p<0.005$] when embedded in the female carrier sentence. This is due to the increased perception of /pɑt/ as /pɒt/ as predicted. More errors were also found for the identification of /pɛt/ [$t= 2.82$, $p<0.01$]. However, in contrast with the predicted perceptual movement towards /pʌt/, the movement was mainly towards /pɜt/. The difference between the carrier sentences in

errors made for /pʌt/ was also significant [t= 2.11, p<0.05], and this was reflected in an increased confusion of /pʌt/ as being /pɒt/ within the female carrier sentence. No significant difference was found between the carriers for /pæt/. For the test words /pɒt/ and /pɜt/ no perceptual changes were expected, and no significant differences were found. There was a relatively large confusion of /pɜt/ as being /pʌt/, however this differed little between the two conditions. The data graphs here and elsewhere show perceptual confusions which occurred more than ten percent of the time within the mixed speaker sentence in comparison to within the control, and bars are one standard error on each side of the mean across the 12 subjects.



b) Effects of the female, the male, and the male imitating female carriers on errors made in the identification of the female test words.

When the male and female carrier were compared, no main effect of carrier sentence was found [F(1,11)= 2.05, p>0.05], but there was a main effect of vowel [F(5,55)= 6.56, p<0.0001] indicating that performance varied across the different test words. There was also an interaction between carrier sentence and vowel [F(5,55)= 3.89, p<0.005]. A comparison between the female and male imitating female carrier sentences found a main effect of carrier sentence [F(1,11)= 7.71, p<0.05], performance being better within the original carrier sentence. There was also a main effect of vowel [F(5,55)= 7.28, p<0.0001], and an interaction [F(5,55)= 6.56, p<0.005], showing that the pattern of errors across the test words for the male carriers was



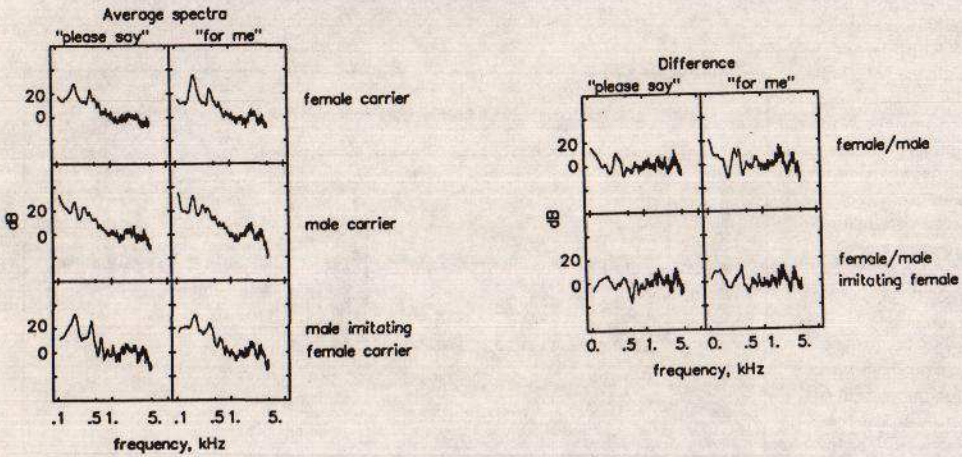
different from the pattern for the original female carrier. Errors in the identification of /pɒt/, /pʌt/ and /pɑt/ were compared between the female and the two other carrier sentences. The comparisons showed that more errors were made in the identification of /pɒt/ when it was embedded in the male carrier [t= 2.54, p<0.05] as well as when it was embedded in the male imitating female carrier [t= 3.56, p<0.005]. This is due to /pɒt/ moving perceptually to /pʌt/ in these conditions, an effect which appears to be slightly larger in the male imitating female condition. No significant difference was found between the female and male carrier for /pʌt/ and /pɑt/. However when the female and male imitating female carriers were compared the difference for /pʌt/ was significant [t= 1.90, p<0.05]. This reflects the apparent perceptual movement of /pʌt/ towards /pɛt/ in the male-imitating condition. For /pæt/, /pɛt/, and /pɜt/ no perceptual movement was predicted, and comparisons between the female carrier and both the male carriers showed no significant differences in errors for these test words. There was a large confusion between /pɜt/ and /pʌt/ in all the conditions.

There is some evidence here to suggest that embedding one speaker's vowel within a second speaker's carrier sentence changes its identity to one appropriate for the second speakers vowel space. This is consistent with an 'extrinsic' speaker normalisation.

4. EXPERIMENT 2.

Putative 'extrinsic' normalisation effects might, in fact, be due to compensation for the carriers' long term spectrum. A central auditory mechanism is supposed to compensate for a sounds' long term spectral characteristics by applying the inverse of the long term average spectrum to subsequent sounds [5, 6, 7]. The carriers used in the first experiment had different long term spectra, thus long term spectra compensation mechanisms may have caused the perceptual vowel movements. Equalising the long term spectrum of the second speaker's carrier sentence with that of the first speaker should therefore eliminate these perceptual movements.

The long term average spectra of the male, the male imitating female, and the female carrier sentences, as well as the difference between the female and male carriers.



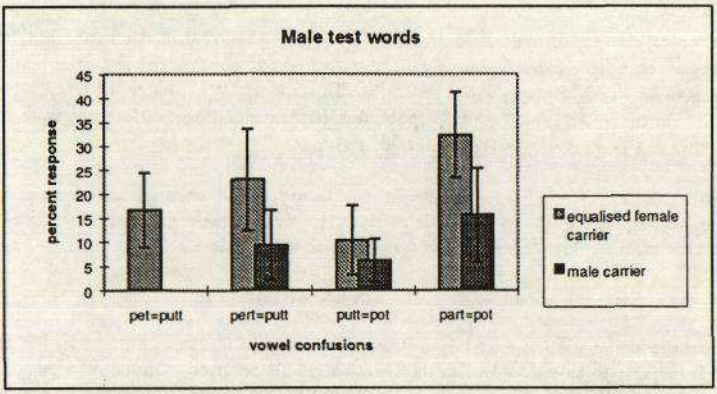
The female and male carriers were filtered by a 'reshaping filter' to give them each others long term spectra

This gave an equalised female carrier (the female carrier sentence with the long term average spectrum of the male carrier), an equalised male carrier (the male carrier sentence with the long term average spectrum of the female carrier), and an equalised male imitating female carrier (the male imitating female carrier with the long term average spectrum of the female carrier). Other aspects of these reshaping filters are as described in [9]. The female test words were presented within the female, the equalised male and the equalised male imitating female carrier sentences, while the male test words were presented within the male and the equalised female carrier sentence. Again each combination was presented 8 times giving 6 vowels × 8 repetitions × 5 carrier with test word combinations = 240 trials for each subject.

Results:

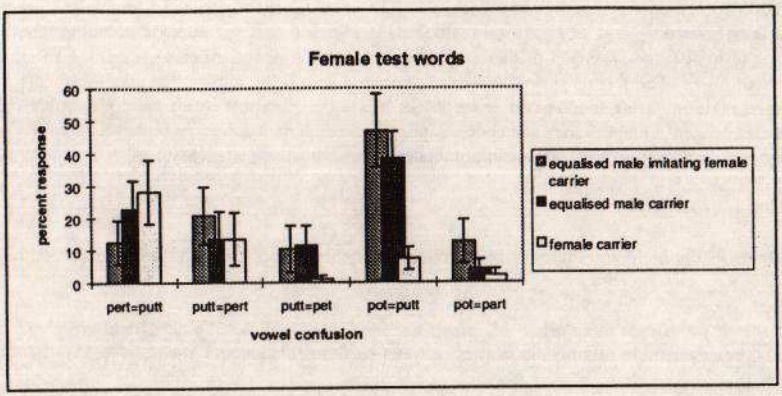
a) Effects of the male and the equalised female carriers on errors made in the identification of the male test words.

There was a significant main effect of carrier sentence [$F(1, 11)= 8.88$, $p<0.01$] and vowel [$F(5, 55)= 3.27$, $p<0.01$], however there was no interaction between carrier sentence and vowel [$F(5,55)= 2.05$, $p>0.05$]. It appears then that although performance within the male carrier was better, and performance varied across the different vowels, the two carriers were not influencing the vowel's perceptual qualities in different ways. Despite the lack of an interaction, vowel confusions that appeared to increase within the equalised female carrier were /pɔt/ being perceived as /pʌt/, /pɛt/ being perceived as /pʌt/, and /pɑt/ being perceived as /pɒt/.



b) Effects of the female, the equalised male and the equalised male imitating female carriers on errors made in the perception of the female test words.

With the male and female carrier there was a main effect of carrier sentence [$F(1,11)= 13.18$, $p<0.005$] and vowel [$F(5,55)= 3.86$, $p<0.005$]. There was also a significant interaction between carrier and vowel [$F(5,55)= 16.73$, $p<0.0001$]. A comparison between the female and equalised male imitating female carrier also found a main effect of carrier sentence [$F(1,11)= 16.54$, $p<0.005$], of vowel [$F(5,55)= 4.84$, $p<0.001$], and an interaction [$F(5,55)= 19.11$, $p=0.0001$]. This pattern of results indicates that more errors occurred within the equalised male and the equalised male imitating female carrier than occurred in the original female carrier. It also indicates that errors varied amongst the different test words. More importantly, the highly significant interactions show that both of the equalised carriers influenced performance across the test words in a different manner to the female carrier. Therefore, equating the long term spectra of the male carriers to that of the female carrier does not eliminate the effects of the carriers on the test words.



More errors were found in the identification of /pɒt/ when embedded in the equalised male carrier sentence [t= 5.52, p<0.0001] as well as in the male-imitating carrier sentence [t= 6.61, p<0.00005]. These results are reflected in a perceptual movement towards /pʌt/, which is increased rather than eliminated in this experiment. The effects of both male carriers were significantly different from the female carrier on the identification of /pʌt/ [t=3.22 and 3.25 respectively, p<0.005]. This appears to be due to a perceptual movement towards /pɛt/. For the vowels /pæt/, /pɛt/, /pɑt/, and /pɜt/ no significant differences were found. Again there was always a large confusion between /pɜt/ and /pʌt/ .

Identification errors in mixed speaker sentences were reduced when the long term spectrum of the female carrier was equated to that of the male carrier. However, these errors were not reduced by equating the long term spectra of the male carriers with that of the female carrier.

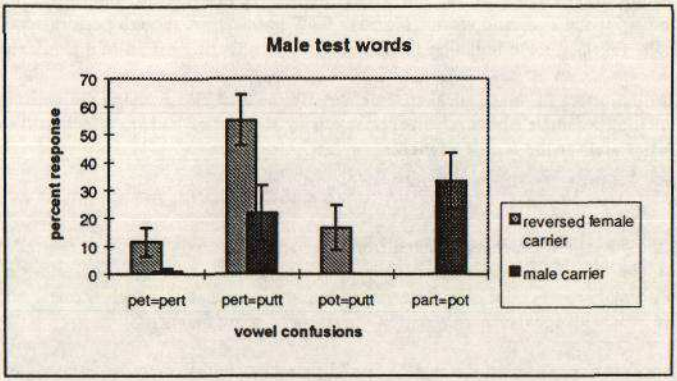
5. EXPERIMENT 3.

If 'extrinsic' compensation for talker differences is due to a mechanism that uses the constraints of speech, then it is likely to be reduced when the speech signal is reversed. On the other hand, if it is due to spectral envelope compensation, carrier reversal should not reduce compensation. This experiment asks whether reversing the carrier sentences influences the perceptual movements observed in the first two experiments. The segments of the carriers both prior to and following the test words were reversed for the female carrier, the male carrier, and the male imitating female carrier. The female test words were presented within the female, the reversed male, and the reversed male imitating female carrier sentence, while the male test words were presented within the male and the reversed female carrier sentence. Again each combination was presented 8 times giving 6 vowels × 8 repetitions × 5 carrier with test word combinations = 240 trials.

Results:

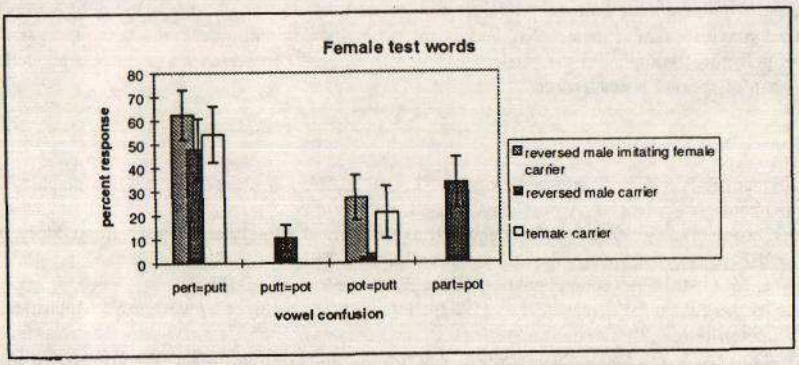
a) Effects of the male and the reversed female carriers on errors made in the identification of the male test words.

There was a main effect of carrier sentence [$F(1,11)$= 6.40, p<0.05] and of vowel [$F(5,55)$= 7.04, p<0.0001]: performance is worse in the reversed condition, and varies across the test words. An interaction [$F(5.55)$= 7.52, p<0.0001] between carrier and vowel indicates that performance differed across the test words in different ways for the different carriers. Assessment of the interaction revealed that the pattern of



vowel confusions with these reversed carriers differed from that found with forward carriers in experiment 1. Significantly fewer errors were made in the identification of /pɑt/ [t= 3.65, p<0.005] when embedded in the reversed female carrier sentence due to a reduction in its confusion with /pɒt/. More errors were made for /pɒt/ [t=2.46, p<0.05], /pæt/ [t=2.55, p<0.05], and /pɜt/ [t=3.11, p<0.005], and no significant difference was found for /pʌt/. The only finding similar to effects with forward carriers was that more errors were found for the identification of /pɛt/ [p<0.05], the movement being mainly towards /pɜt/.

b) Effects of the female, the reversed male, and the reversed male imitating female carriers, on errors made in the perception of the female test words.



When the reversed male and the female carrier were compared there was no main effect of carrier sentence [F(1,11)=4.34, p>0.05]. A main effect of vowel [F(5,55)=11.36, p<0.0001] indicated that performance varied across the test words. There was also an interaction between carrier sentence and vowel [F(5,55)=6.11, p<0.0001]. The reversed male carrier was influencing errors made in the test words in a different way to the female carrier. Significant differences were found for /pʌt/ [t= 2.03, p<0.05] and /pɑt/ [t= 3.71, p<0.005]. Also it was found that significantly fewer errors were being made in the identification of /pɒt/ when embedded in the reversed male carrier sentence [t= 2.35, p<0.05], due to a reduction in /pɒt/'s confusion with /pʌt/. No significant differences were found for the vowels /pæt/, /pɛt/, and /pɜt/. There was again a relatively large confusion between /pʌt/ and /pɜt/ in all the conditions. None of the vowel movements predicted by speaker normalisation occurred in this reversed carrier experiment. A comparison between the female and reversed male imitating female carrier sentences found a main effect of

carrier sentence [$F(1,11)= 8.30$, $p<0.05$], with more errors being made in the reversed condition. There was also a main effect of vowel [$F(5,55)=12.78$, $p<0.0001$], however no significant interaction was found between carrier sentence and vowel [$F(5,55)= 0.48$, $p>0.05$]. Although performance differed among the test words, the reversed male imitating female carrier was effecting errors in a similar way to the female carrier.

Identification errors in mixed speaker sentences are reduced by reversing the carrier when the carrier is a male imitating a female's pitch. However, reversing the carrier had no effect on these errors for the normal male carrier and for the female carrier.

6. DISCUSSION.

Not all the vowel movements predicted from overlaps on the vowel space occurred when the male and female speaker's vowels were embedded in each others sentences. This indicates that factors other than speaker normalisation could have been generating the errors. Nevertheless, most of the confusions observed are consistent with speaker normalisation operating to some extent.

If this speaker normalisation is actually caused by a compensation for spectral envelope distortion, then equalising the long term averaged spectra of the carrier sentences should reduce the confusion. This did occur in experiment 2 when the female carrier was equalised to the male carrier and presented with the male speaker's test words. However, when the male and the male imitating female carriers were equalised to the female carrier, perceptual confusions remained. Therefore, some other form of speaker normalisation could have been operating in these conditions.

If this speaker normalisation is brought about by a phonetic mechanism, then the confusions that it causes should be reduced when the carriers are reversed. This did reduce confusions in experiment 3, but only for the carrier where the male imitated the females voice-pitch. Interestingly, this is the only carrier that produced mixed speaker sentences that were heard as wholly originating from the same speaker, so some form of perceptual grouping of the carrier and test words might be necessary to engage a phonetic mechanism of speaker normalisation.

7. REFERENCES

[1] Assman, P. F., Neary, T. M., and Hogan, J. T. (1982). "Vowel identification: Orthographic, perceptual and acoustic aspects," J. Acoust. Soc. Am. 71, 975-989.

[2] Ainsworth, W.A. (1975). "Intrinsic and extrinsic factors in vowel judgements," in Auditory Analysis and Perception of Speech, edited by G. Fant and M. Tatham (Academic, London), pp. 103-133.

[3] Joos, M. (1948). "Acoustic phonetics," Lang. 24 (Supplement), 1-136.

[4] Ladefoged, P., and Broadbent, D E. (1957). "Information conveyed by vowels," J. Acoust. Soc. Am. 29, 98- 104.

[5] Watkins, A. J. and Makin, S. J. (1994). "Perceptual compensation for speaker differences and for spectral- envelope distortion," J. Acoust. Soc. Am., In press.

[6] Watkins, A. J. (1988). "Spectral transitions and perceptual compensation for effects of transmission channels," in Proceedings of the 7th Symposium of the Federation of Acoustical Societies of Europe: Speech '88. edited by W. Ainsworth and J. Holmes (Institute of Acoustics, Edinburgh.).

[7] Watkins, A. J. (1991). "Central auditory mechanisms of perceptual compensation for effects of transmission channels," J. Acoust. Soc. Am. 90, 2942-2955.

[8] Dechovitz, D. (1977). "Information conveyed by vowels: A confirmation," Haskins Laboratories: Status report on speech research SR-51/52, 213-219.

[9] Van Bergem, D. R., Pols, L. C. W., and Koopmans-van Beinum, F. J. (1988). "Perceptual normalization of the vowels of a man and a child in various contexts," Speech Communication 7, 1-20.