Proceedings of The Institute of Acoustics

INFLUENCE OF TRANSMISSION CHANNEL CHARACTERISTICS ON PERCEIVED VOWEL QUALITY

A.J.Watkins (1), Helen Pattison (2) and C.J.Darwin (2)

(1) Department of Psychology, University of Reading, Reading RG6 2AL
(2) Experimental Psychology, University of Sussex, Brighton BN1 9QG

ABSTRACT

Three experiments address the question of how the perception of F1 is influenced by passing isolated steady-state vowels (Expt 1) or /jVj/ syllables (Expt 2) through filters which change the relative amplitudes of harmonics in the F1 region. The results of the experiment show reliable changes in the /I/-/E/ phoneme boundary that move in opposite directions depending on whether the relative amplitudes of the F1 harmonics are changed or whether the whole of the F1 region is simply reduced in amplitude. The third experiment introduces the steady-state vowels with a natural carrier phrase that is either unfiltered or has been filtered in the same way as the vowels. This experiment failed to show any phoneme boundary shift or any difference between the different filterings of the carrier phrase.

INTRODUCTION

The well-known studies of Miller and Nicely (1955) and of Pickett (1957) demonstrated how the perception of particular features of consonants and vowels can be impaired or lost when speech has been filtered or noise added. They indicate which frequency regions must be preserved for the listener to be able to make particular phonetic distinctions. Less drastic perturbations of the speech spectrum, such as giving it pre-emphasis, cause rather little perceptual distress, although they can frustrate template matching schemes for assessing vowel colour (Klatt, 1982a).

The frequency of a spectral peak will be little affected by gross changes to spectral balance and so provides an ecologically sensible primitive for coding phonetic quality. It is possible that human listeners code phonetic quality by the frequencies of formant peaks rather than by the overall shape of the spectrum since their phonetic judgements are generally so much more influenced by changes in formant frequencies than by changes in formant amplitudes or bandwidths or by the presence of spectral zeroes (Klatt, 1982b).

But if a sound's spectrum is coded simply in terms of the frequency of its auditorially resolved peaks, important phonetic information is lost, especially in the region of the first formant. The auditory system's frequency resolution (Moore and Glasberg, 1983) allows individual harmonics to be separated in the F1 region, but generally not (at least for low-pitched voices) in the higher formants. Consequently, a scheme that represented <u>only</u> the frequency of spectral peaks would not capture the position of F1, since it is defined by the envelope of the resolved harmonics. The position of F1 must be computed from the amplitudes of the separate harmonics in the region of the formant peak (Darwin and Gardner, 1985).

Transmission channel and vowel quality

The experiments presented here examine (1) how the perception of the phonetic quality of vowels along a continuum that differs in F1 is influenced by filtering them through a transmission channel that changes the relative amplitudes of harmonics in the F1 region and (2) how the human listener does or does not compensate for such distortions.

The experimental paradigm used measured changes in the phoneme boundary for an /I/-/E/ vowel continuum differing in F1 when the vowel sounds were filtered in various ways in the F1 region. Changes in the perceived F1 lead to shifts in the phoneme boundary. A higher perceived F1 gives a lower boundary.

EXPERIMENT 1

Method

In the first experiment, a continuum of vowels of 56 ms duration with 16ms rise/fall times was synthesized on a fundamental of 125 Hz. The sounds varied in the value of F1 from 417 Hz to 543 Hz in equal 21 Hz increments, giving /I/ sounds at low F1 values and /E/ sounds at high F1's. The values of the second to fifth formants were 2.3 kHz, 2.9 kHz 3.8 kHz and 4.6 kHz respectively. The bandwidths of the first three formants were 90 Hz, 110 Hz and 170 Hz.

This set of normal vowels was filtered through five FIR filters with varying slopes of attenuation in the the first formant region between 1kHz and 250 Hz. The maximum attenuation was 6, 12, 18, 24 or 30 dB at 250 Hz. As a control for the effect of an overall lowering of amplitude of the first formant, the vowels were also passed through filters with flat attenuation below 1 kHz of 3, 6, 9, 12 and 15 dB, stepping up to 0 dB attenuation at 1.5 kHz. The sloping and flat filters are illustrated in Figure 1. Spectra of a vowel from the middle of the continuum and filtered through the two most extreme filters are shown in Figure 2.

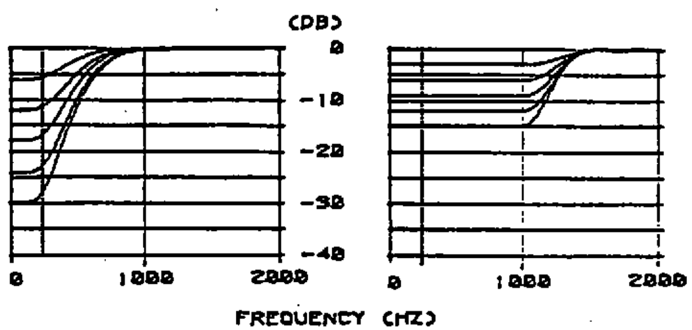

Figure 1. Transfer functions for sloping (left) and step (right) filters

Transmission channel and vowel quality

Procedure

Each member of a continuum was repeated 10 times giving a total of 70 stimuli per condition. The 700 stimuli were presented in random order. 12 normally hearing subjects listened to the sounds over headphones in a sound-proof cubicle. Sounds were produced on-line from a VAX-11/780 computer via an LPA-11K at a sampling rate of 10 kHz, and low-pass filtered at 4.5 kHz and 48dB/oct. Subjects responded on a VDU keyboard, indicating whether they thought the sound was more like /I/ or /E/ by pressing the "i" or "e" key. The experimental trials were preceded by 70 practice trials.

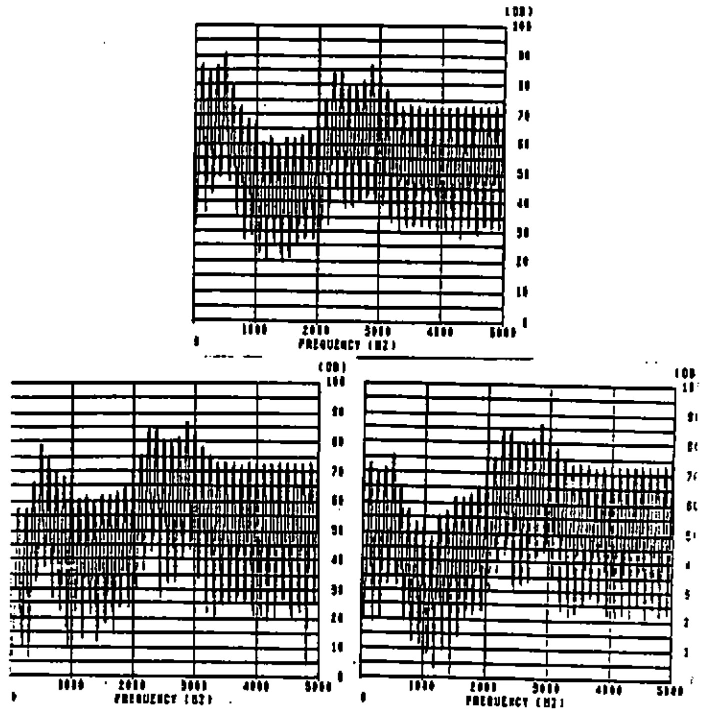

Figure 2. Vowel spectra from original (top), sloping (left) and step (right) continua.

Results

Phoneme boundaries were calculated for each subject for each continuum by probit analysis. The results are illustrated in Fig.3. Putting different sloping filters on the vowel sounds caused a downward shift in the F1 boundary, thus F1 is perceived as progressively higher than it is in the non-filtered sounds. This shift is in the expected direction, and of an appropriate size on the assumption that the F1 value is taken as the peak in the local harmonic envelope. When the F1 intensity was uniformly lowered, there was an unexpected but large shift in the opposite direction. A lower amplitude of F1 is perceived as equivalent to a lower F1 frequency. An analysis of variance on the difference between the phoneme boundaries in the unfiltered condition and the filtered conditions revealed a main effect of filter shape ($F[1,11]=109.33$; $p<0.0001$) and an interaction between filter shape and level of attenuation ($F[4,44]=28.43$; $p<0.0001$).

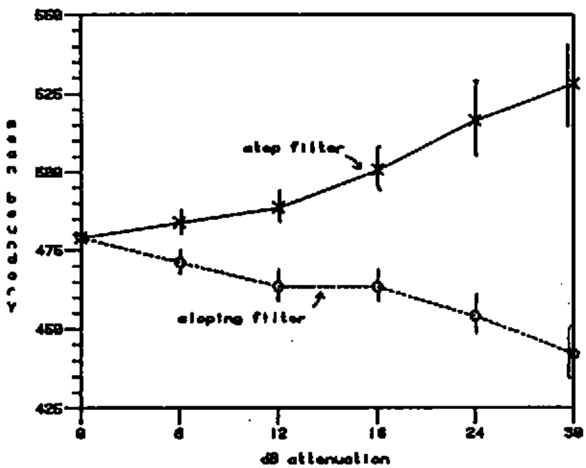

Figure 3 Mean phoneme boundaries from Experiment 1.

Discussion

The experiment shows that when the filter characteristics of the transmission channel change unpredictably, listeners are, not surprisingly, unable to compensate, and perceive an altered vowel quality.

EXPERIMENT 2

The second experiment investigated whether the vowel shift remains when vowels are put in a /jVj/ context. We anticipated that the presence of formant transitions in the F1 region might help the listener to disentangle vocal tract and transmission channel effects.

Transmission channel and vowel quality

Method

Two continua were synthesized which placed an /I/-/E/ F1 continuum in a /jVj/ context. One continuum (F1 250) had F1 transitions that started at 250 Hz and rose in 100ms to the vowel's steady state. The steady state lasted 56ms and then F1 fell in 100ms to 250 Hz. The other (F1-160) had F1 transitions starting 160 Hz below the steady state F1 value. F1 values of the vowel's steady state varied between 330 Hz and 459 Hz in 21 Hz increments. 8 continua were then made from these 2, by filtering them through 4 filters: 2 sloping filters with maximum attenuation of 12 dB and 24 dB, and 2 step filters with attenuation of 6 dB and 12 dB. 8 normally hearing subjects took part in the experiment, and the experimental procedure was similar to that described for Experiment 1.

Results

The results shown in Fig.4 are not fundamentally different from those of Experiment 1. The boundary still shifts downwards when the /jVj/ syllables are filtered through a sloping filter and upwards when filtered through a step filter. An analysis on the differences between the filtered and unfiltered conditions revealed a main effect of filter shape ($F[1,7]=23.95$; $p<0.002$) and an interaction between filter shape and level of attenuation ($F[1,7]=10.49$; $p<0.01$), but no significant effect of the different starting point of the F1 transitions.

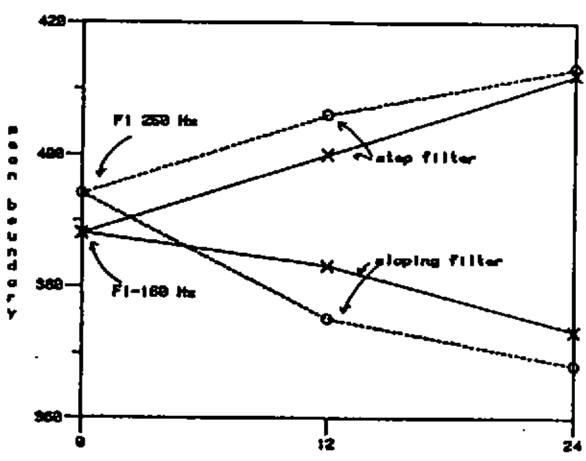

Figure 4. Mean phoneme boundaries from Experiment 2.

Discussion

The second experiment has shown that the presence of formant transitions in the region of the sloping filter does not undo the perceptual change that occurs as a result of the unpredictable transmission channel characteristics.

EXPERIMENT 3

Ladefoged and Broadbent (1957) showed that the perception of vowel quality is influenced by the F1 and F2 values of other vowels occurring in the same auditory context. Experiment 3 investigated whether the perception of vowel quality of filtered vowels is influenced by placing the vowel in an auditory context which has been similarly filtered. Since the context, a natural spoken sentence, remained constant except for the filtering, subjects should, in principle, be able to figure out the characteristics of the transmission channel.

Method

This experiment used the original vowel continuum from Experiment 1 filtered using sloping filters with either 6 dB, 12 dB or 24 dB attenuation. To provide an auditory context, an introductory sentence 'Please say what this word is.' was eloquently spoken by CJD. The vowel sounds were then placed 300ms after the introductory sentence which was either unfiltered or filtered in the same way as the vowels. This gave 7 conditions in total: 1 in which the introduction and vowels were unfiltered, 3 in which the vowels were filtered with one of the 3 filters but the introduction was unfiltered, and 3 in which both the introduction and vowels were filtered in the same way.

Twelve subjects took part in this experiment, and the procedure was similar to that used in the previous two experiments.

Results

The results (Fig 5) show no shift of the phoneme boundary in any condition. The expected replication of experiments 1 and 2 when the vowel was preceded by the unfiltered sentence did not occur. Consequently, the absence of a shift, when the vowel was preceded by a similarly filtered sentence, is hard to interpret. An analysis of variance revealed no significant difference between the unfiltered condition and any of the filtered conditions, and no significant difference between the conditions in which the whole stimulus was filtered and those where only the vowel sound was filtered.

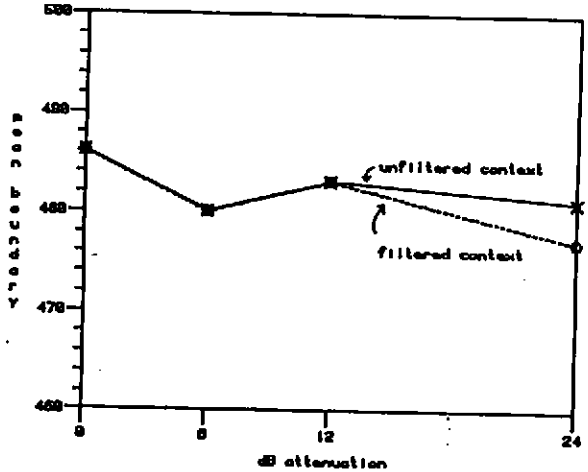

Figure 5. Mean phoneme boundaries from Experiment 3.

Summary And General Discussion

The first two experiments reported here successfully demonstrated a change in phoneme boundry along an /I/-/E/ F1 continuum when vowels are unpredictably filtered. The effect shown by steady-state vowels in the first experiment is replicated with vowels in a /jVj/ syllabic context in the second experiment. However, when the isolated synthetic vowels are preceded by a natural carrier phrase, the shift in phoneme boundary is abolished, whether the carrier phrase remains unfiltered or is filtered in the same way as the vowel. It is not clear why the last experiment gave such results al;though it is possible that the natural carrier phrase made it hard for subjects to maintain a stable decision criterion for the phoneme boundary.

The phoneme boundary shift that we found in the first two experiments provides a potentially useful paradigm for investigating how differences in transmission functions influence the perception of phonetic categories.

ACKNOWLEDGMENTS

This research was supported by grants from the SERC. The first experiment was run by David Kirby as an undergraduate project at Sussex.

REFERENCES

C.J.Darwin and R.B.Gardner (1985) 'Which harmonics contribute to the estimation of first formant frequency.' Speech Communication, 4, 231-235.

D.H.Klatt (1982a) 'Speech processing strategies based on auditory models.' In R.Carlson and B.Granstrom (Eds) The representation of speech in the peripheral auditory system. Elsevier Biomedical Press.

D.H.Klatt (1982b) 'Prediction of perceived phonetic distance from critical-band spectra: a first step.' Proc IEEE International Conference on Speech, Acoustics and Signal Processing, 1278-1281.

P.Ladefoged, and D.E Broadbent, (1957) 'Information conveyed by vowels', J.A.S.A.,29,98-104.

G.A.Miller and P.E.Nicely (1955) 'An analysis of perceptual confusions among some English consonants', J.A.S.A., 27, 338-352.

B.C.J.Moore and B.R.Glasberg (1983) 'Suggested formulae for calculating auditory-filter bandwidths and excitation patterns.' J.A.S.A., 74, 750-753.

J.M.Pickett (1957) 'Perception of vowels heard in noise of various spectra', J.A.S.A, 29, 613-620.