INDIVIDUAL VARIABILITY IN THE PERCEPTUAL WEIGHTING OF CUES TO STOP PLACE AND VOICING CONTRASTS

V. Hazan and B. Shi
Dept of Phonetics and Linguistics, University College London, UK

ABSTRACT

Listener variability in the perceptual weighting of acoustic cues to stop contrasts has been studied in a group of 50 listeners. For the place contrast, perceptual weighting given to the burst and transition cues varied widely with listeners and vocalic contexts. More homogeneous results were obtained for the voicing contrast. It is argued that listeners vary in their perceptual strategies and that acoustic cues vary in terms of their robustness.

INTRODUCTION

Phonetic contrasts are marked by a multiplicity of acoustic cues. The relative weighting given to cues to contrasts in manner, voicing and place of articulation has been shown to vary according to vocalic context [1,2] and speaker characteristics [1]. Although the emphasis in studies of cue weighting has been on the presentation of averaged results, many studies have found that some individual listeners may show quite different cue weighting strategies than the norm.

The aim of this study was therefore to quantify the amount of variability in cue weighting seen in a large and relatively homogeneous listener population. In order to evaluate the effect of contrast type and vocalic context on individual variability, listeners were tested on two different contrasts, each presented in three vocalic contexts.

METHODOLOGY

Stimuli

Natural tokens produced by a male English speaker were used as a base for copy-syntheses which were obtained using a version of the Klatt software synthesizer produced by Sensimetrics (KLSYN88). Once a close copy of the minimal pair was obtained, a set of continua was prepared by interpolating the parameters under investigation.

/g/-/d/ place contrast

The acoustic cues under investigation were the burst transient and F2/F3 transitions at vowel onset. In order to evaluate the effect of vocalic context on cue-weighting, three minimal pairs were used: GATE-DATE, GAT-DAT, GEET-DEET. The initial burst was synthesized through the parallel branch of the synthesizer, by exciting five formants with noise for 5 ms. The formant values for F2 and F5 were fixed at 1800 Hz and 4000 Hz. F3 varied from 2300 Hz at the /g/ endpoint to 4200 Hz at the /d/ endpoint. F4 varied from 2700 Hz at the /g/ endpoint to 4600 Hz at the /d/ endpoint. The amplitude of the two formants varied from 75 dB at the /g/ endpoint to 60 dB at the /d/ endpoint. The burst was identical for all three contrasts.

The vowel was synthesized through the cascade branch of the synthesizer. Formant transitions of F2 and F3, which were matched to values measured in the natural tokens and which extended over the first 50 ms of the vowel, constituted the second cue to the place contrast. A full description of the stimuli is given in [3].

In order to evaluate the relative contribution of each cue to the perception of the contrasts, three test conditions were prepared for each minimal pair. In the "Combined-cue" condition, both cues were varied together; in the "No transitions cue" condition, the formant transitions were fixed at a neutral value; in the "No burst cue" condition, the burst was removed through waveform editing.

/g/-/k/ voicing contrast

The two main cues to the contrast under investigation were voice onset time (VOT), i.e. the duration between burst release and voicing onset, and F1 onset frequency. The minimal pairs were: GATE-KATE, GAT-KAT and GEET-KEET. VOT varied from 15 ms to 115 ms following burst onset, in 10 ms steps. As VOT increased, so did the cutback in F1 relative to higher formants that were present within the aspiration portion.

The second cue under investigation was F1 onset frequency. The F1 transition took place over the first 50 ms of the vowel. At the /g/ endpoint, for the GATE-KATE contrast, F1 onset frequency was set at 327 Hz rising to 500 Hz. For the GAT-KAT contrast, F1 onset frequency was set at 420 Hz rising to 630 Hz. For the GEET-KEET contrast, F1 onset frequency changed from 310 Hz to 290 Hz.

For each minimal pair, two stimulus conditions were prepared: (1) "Combined-cue", in which both VOT and F1 onset frequency were co-varying, and (2) "No F1 transition cue", in which F1 at onset was fixed at the frequency reached at the end of the formant transition period.

Listeners

Listeners were 50 volunteers with pure tone thresholds of 20 dB HL or better from 0.25 to 8 kHz in both ears. The listeners, students at UCL, ranged in age from 18 to 32 years (mean: 21.1 years, s.d. 3.09), and were native English speakers who had no training in phonetics and little or no previous exposure to synthetic speech.

Test procedure

The identification test procedure was computer-controlled. Stimuli down sampled at 10 kHz were presented as two alternative forced choice identification tests. Stimuli were amplified to a comfortable listening level and presented through AKG 240 DF headphones. Listeners responded to each stimulus by pressing a touch sensitive response box.

For each contrast, all test conditions were randomised together in order to reduce range effects. 32 responses per stimulus were collected over four sessions for each listener.

RESULTS

A statistical approach based on Generalized Linear Models (GLMs) was used to determine the extent to which the change in deviance between the combined-cue condition and each of the single-cue conditions for a given contrast was significant. This technique, analogous to ANOVA, was used as it is especially tailored to the analysis of multi-variate data involving binary responses (for a full description, see [2]).

On the basis of previous results [2], it was hypothesized that, for the place contrast, there would be evidence of listener groups showing different usage of the burst and formant transition information. For each contrast, the percentage of listeners showing significant differences in identification function between the combined-cue and each single-cue condition was calculated (see Table 1). Some listeners were affected by the removal of either cue. However, it can also be seen that for each minimal pair in the place contrast tests, some listeners were unaffected by the removal of the burst information, while others were unaffected by the removal of the formant transition

information. A small number of listeners (12% for GATE-DATE, 6% for GAT-DAT, 8% for GEET-DEET) were unaffected by the removal of either cue and could therefore reliably label the contrast on the basis of whatever cue information was present.

Table 1. Percentage of listeners showing significant deviances in single-cue conditions of the place contrasts relative to the combined-cue condition.

	No burst cue	No trans. cue
GATE-DATE	68 %	70 %
GAT-DAT	84 %	42 %
GEET-DEET	88 %	10 %

The number of listeners significantly affected by the removal of the formant transition cue varied widely across vocalic contrasts from 10% for the GEET-DEET contrast to 70% for the GATE-DATE contrast. The percentage of listeners significantly affected by the removal of the burst cue varied between 68% (GATE-DATE contrast) and 88% (GEET-DEET contrast).

Table 2. Percentage of listeners showing significant deviances in the single-cue condition of the voicing contrasts relative to the combined-cue condition.

	F1 transition removed
GATE-KATE	2 %
GAT-KAT	34 %
GEET-KEET	4 %

For the /g/-/k/ contrast, an examination of individual results reveals less variability than for the place contrast. The removal of the F1 onset cue had little effect for the GATE-KATE and GEET-KEET contrasts. However, 34% of listeners showed a significant effect of removal of F1 onset cue for the GAT-KAT contrast.

DISCUSSION

The results of this study gives some evidence of the extent of variability in cue-weighting across contrasts and vocalic context. For the /g/-/d/ place contrast, the perceptual weighting of the burst and formant transition cues varied quite considerably according to vocalic context. For the GATE-DATE contrast, a similar number of listeners were affected by the removal of either cue. For the GEET-DEET contrast, there was a clear dominance of the burst cue over the formant transition cue. For the GAT-DAT contrast, a less extreme imbalance was obtained, with 84% of listeners affected by burst removal vs 42% by the removal of the formant transition cue.

The effect of vocalic context on cue-weighting appeared to be related to the degree of acoustic prominence of the cue. For example, the greatest effect of the removal of F1 transition for the voicing contrast was found for the vowel environment with the highest first formant and therefore the greatest transition extent. Similarly, the least effect of F2 transition removal in the place contrast was obtained in the context of /i/ in which the formant transitions are less pronounced due to the high F2 of the vowel.

Within a specific vowel environment, the evidence of clear individual differences in the perceptual effect of acoustic cue removal would suggest that listeners do differ in the use that they make of acoustic cue information contained in the speech signal. This confirms results of previous studies involving nonsense syllables [4] and identification tests for speech contrasts [2] and goes some way towards explaining some contradictory results found in the literature on the perceptual

weighting of acoustic cues to speech contrasts.

At the speech pattern processing level, the fact that the effect of cue-weighting was more variable for the place contrast than for the voicing contrast suggests that cues differ in terms of their "robustness". For the voicing contrast, VOT is clearly dominant cue for a vast majority of listeners, whereas for the place contrast, relative importance of burst and formant information varies greatly across listeners and across vocalic contexts. A better understanding of which acoustic cues are least subject to listener and contextual variability has implications for work on cue-enhancement in synthesized and degraded natural speech.

It may be hypothesized that individual variability in the use of acoustic cues is due not to audiological differences but to the development of different perceptual strategies during language acquisition, where individuals may focus on one of several redundant cues contained in the speech signal. There is ample evidence of individual differences in language and speech development (for a review, see [5]). Some evidence of individual differences in the perceptual weighting given to cues for a "bees/peas" voicing contrast was seen in a study [6] in which only 60% of 4-year old children were affected by a change in the vowel stem, which introduced conflicting spectral cues. This was similar to the percentage of adults affected by the conflicting cue.

The presence of sizable variability and of different perceptual strategies within a homogeneous population of listeners highlights the importance of considering the effect of human factors in the interpretation of the results of perceptual experiments. Indeed, the particular composition of the listener group might have a strong effect on scores obtained, especially if the listener group is small. The existence of individual differences in perceptual strategies might also go some way towards explaining the great difference in performance seen in the use of speech processing aids by deafened adults fitted with cochlear implants, for example.

ACKNOWLEDGEMENT

This work was funded by a project grant from the Science and Engineering Research Council and the Ministry of Defence (GR/F 33735).

REFERENCES

[1] Dorman, M.F., Studdert-Kennedy, M., Raphael, L.J. (1977), "Stop-consonant recognition: release bursts and formant transitions as functionally-equivalent, context-dependent cues", *Perception and Psychophysics*, vol. 22, pp. 109-122.

[2] Hazan, V. & Rosen, S. (1991), "Individual variability in the perception of cues to place contrasts in initial stops", *Percept. Psychophysics*, vol. 49, pp. 187-200.

[3] Hazan, V. and Shi, B. (1993), "Individual variability in the identification of plosive place and voicing contrasts" *Speech, Hearing and Language: UCL Work in Progress*, vol. 7, pp. 77-94.

[4] Santi, S. and Grenié, M. (1990), "Individual strategies in synthetic speech evaluation", *Proceedings of the ESCA workshop on Speech Synthesis*. Autrans, France, September 1990, pp. 265-68.

[5] Bates, E., Bretherton, I. and Snyder, L. (1988), *From first words to grammar: individual differences and dissociable mechanisms*, Cambridge: Cambridge University Press.

[6] Howell, P., Rosen, S., Lang, H. and Sackin, S. (1992), "The role of F_1 transitions in the perception of voicing in initial plosives", *Speech, Hearing and Language, Work in Progress UCL*, vol. 6, pp. 117-126.