ACOUSTICAL CUES IN /R-W/ DISCRIMINATION

Laurie K. Fitzgerald & Elzbieta B. Slawinski
University of Calgary, Department of Psychology

1. Introduction

There has been a great deal of controversy concerning the categorical perception of the semivowels /r/ and /w/. Although research has clearly shown that the above two phonemes are perceptually differentiated from one another on the basis of primarily two acoustical features, namely the onset frequency and transition duration of their respective second (F2) and third (F3) formants, the same has not elucidated, however, the "primacy" or perceptual "weight" of either of these two cues. That is to say, it has not been definitively specified which feature represents the more salient cue underlying the /r-w/ phonemic distinction. Consequently, when ever there has been discrepant results in the /r-w/ perceptual literature, an explanation which often surfaces relates to the acoustical characteristics of the stimuli used in the studies. Discordant results are commonly attributed to differences in the acoustical properties of the stimuli employed. In some studies, the stimuli utilized contained only formant onset frequency information whereas in other studies, the stimuli contained only the formant transiton information. In an attempt to better understand the role the two cues play in the phonemic perception of /r/ and /w/, the present study was undertaken. It was hoped that by comparing identification performances on continua with formant transition information alone, formant onset frequency information alone, and then a combination of both types of information, some light may be shed on the above issue. Working from the premise that the preceding speculation was correct, it was hypothesized that a main effect for continuum type would be exhibited.

2. Method

Subjects: The subjects were 35 male and female adults between the ages of 18 and 35. All were native speakers of English, and all had normal peripheral hearing.

Stimuli: Three 7-step male adult /r-w/ continua were synthesized via the Klatt Cascade/Parallel Software Program (Klatt, 1980). The stimuli in the continua corresponded to the words rock and walk and were 330 msec in duration. The parameter values used in the synthesis of the three continua's endpoint stimuli were patterned after Klatt(1980) guidelines and those obtained from an acoustical analysis of an adult males' recorded versions of rock and walk. The parameter values of the vowel nucleus /a/ and the stop /k/ were held constant across the continua. The stimuli were also matched for pitch, intonation, and amplitude contour. The acoustic tokens in the three continua did vary, however, according to F2 and F3 onset frequencies and F2 and F3 transitions durations. In the first continuum (C1), the acoustic tokens varied with respect to both F2 and F3 onset frequencies and F2 and F3 transition durations. F2 onset frequency varied from 1070 Hz in the endpoint stimulus corresponding to a clear token of the word rock to 650 Hz in the endpoint stimulus corresponding to a clear token of the word walk. F3 onset frequency varied in the same manner from 1460 Hz to 2150 Hz. F2 transition duration varied from 15 msec in the endpoint stimulus corresponding to a clear token of rock to 75 msec in the endpoint stimulus corresponding to a clear token of walk. Similarly, the transition duration of F3 varied from 75 msec to 15 msec. In the second continuum (C2), the acoustic tokens varied only with respect to F2 and F3 onset frequencies, and the durations of the second and third formant transitions were kept constant at a value of 45 msec. In the third continuum (C3), the acoustic tokens varied only with respect to F2 and F3 transition durations, and the onset frequencies of the second and third formants were kept constant at 860 Hz and 1805 Hz respectively.

Upon completion of synthesis, two sets of 35 trials were prepared for each continua, in which the 7 stimuli in each respective continuum was presented 5 X each in random order.

Procedure: The testing was done on an individual basis and was conducted in the Audition and Speech Laboratory at the University of Calgary. Prior to testing, subjects underwent audiometric screening. Normal performance was defined as thresholds 20 dB or lower for the frequencies of 1000, 2000, and 4000 Hz.

The test proper consisted of a two-alternative, forced-choice identification task embedded in a pseudo video-game. The video-game was played on a Apple II MacIntosh computer and involved listening to the stimuli of the continua over a loudspeaker and then identifying whether the sound heard was rock or walk by choosing a corresponding picture within the video-game.

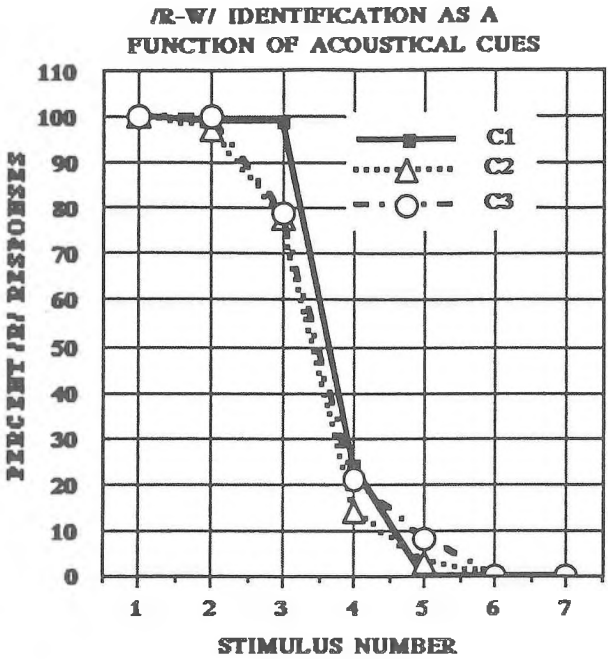
Each subject performed an identification task with each continuum type, with the trial sets and task order counterbalanced across subjects. All three identification tasks were completed in one sitting, with a 5 minute break between the running of the second and third task. A formal training protocol was not implemented in the study, but as a means of acquainting the subjects with the nature of the task and Mac controls, 4 "dry-runs" of the video-game involving randomized endpoint stimuli were conducted with each of the three identification tasks. On average, testing took twenty minutes to complete.

3. Results

All inferential statistics were computed with the Statview Software Program on an Apple II Macintosh computer.

The mean percentages of /r/ responses were computed for all three continua and plotted as a function of stimulus number. The identification function of the two-cue continuuum was different than that evidenced in either of the two one-cue continua. Although in all three continua stimuli 1-4 were identified as /r/, as evidenced by the high percentage of /r/ responses for those stimuli, and conversely, stimuli 5-7 as /w/, as evidenced by the low percentage of /r/ responses for those stimuli, the percentages of /r/ responses for stimulus 3 in the two, one-cue continua were lower than that exhibited in the one, two-cue continuum. Stimulus 3 had a 99% identification rate in C1, the continuum containing both onset frequency and transition duration formant information. In contrast, the continua containing information soley on onset frequency (C2) or transition duration (C3) had response rates of only 78 and 79% respectively.

In addition to characterizing the subjects perceptual categorizations as a function of stimulus number and /r/ response percentage, phonemic boundaries, which mark the 50% point between sound categories, were calculated for each individual for each continuum. This was done by means of interpolating the stimulus number at the 50% crossover on the unsmoothed response functions. Mean category boundaries for each continuum were obtained by averaging the individual



category boundaries. The mean phonemic boundary values were 3.73, 3.41, and 3.54 for C1, C2, and C3 respectively. A repeated measures analysis of variance on the factor of continuum type revealed the latter to be signficant ($F(2,68) = 8.568$; $p < .0005$). The mean phonemic boundary value corresponding to C1 was significantly different from the mean phonemic boundaries evidenced in C2 (Scheffe $F=8.483$; $p < .05$) and C3 (Fisher PLSD=.158; $p<.05$). The mean phonemic boundaries of C2 and C3, however, were not significantly different from one another.

4. Conclusions

(1) The finding that perceptual categorization of /r/ and /w/ varied as a function of two cue vs one cue availability implies that caution should be exercised when comparing data obtained on /r/ and /w/ stimuli incorporating only one of the differentiating salient acoustical dimensions vs incorporating both. This finding also suggests that the two cues are perceived in relation to one another. In turn, this provides support for the theory espoused by Wood (1975) and Soli (1980) that perception of speech sounds does not precede directly and independently from acoustical cues but rather is mediated by higher order coding levels which incorporate dependencies among acoustical cues. It is recognized, however, that the present study's support for the above theory is limited given the nature of the experimental task employed. A more stringent test would involve a trading relations design as described below in point (4).

(2) The finding that perceptual categorization of /r/ and /w/ did not vary as a function of which individual cue, formant onset frequencies or formant transition durations, was available for the same, suggests that the latter are perceptually equivalent as discriminative cues for /r/ and /w/.

(3) Because the data was collapsed across individuals, it was not possible to determine if the two cues were perceptually equivalent at the level of the individual. It would be interesting to investigate the possible existence of different individual processing strategies and/or perceptual biases brought to play in the phonemic perception of /r/ and /w/. Research conducted on stop consonant discrimination suggests that individual perceptual biases toward spectral vs temporal cues do in fact occur (Fourcin 1988). Furthermore, it appears that the perceptual biases eliminate the discriminative advantage of two-cue availability.

(4) Further research could involve a more rigorous test of the discriminative equivalence of the above two acoustical properties via a trading relations experiment. In such a study, the two acoustical dimensions could be combined in such a way as to cue the opposing phoneme ie. one dimension's parameters would be typical of /r/ and the other's parameters would be typical of /w/. Perceptual equivalence of the two cues would be evidenced by a 50% indentification rate to all of the stimuli within the continuum.

5. References

Fourcin, A. (1988). Links between voice pattern perception and production. In B.A.G. Elsendorn & H. Bouma (Eds.), Working models of human perception (pp. 67-91). Toronto: Academic Press.

Klatt, D.H. (1980). Software for a cascade/parallel formant synthesizer. The Journal of the Acoustical Society of America, 67 (3), 971-995.

Soli, S.D. (1980). Some effects of acoustic attributes of speech on the processing of phonetic feature information. Journal of Experimental Psychology: Human Perception and Performance, 6, 622-638

Wood, C.C. (1975). A normative model for redundancy gains in speeded classifications: Applications to auditory and phonetic dimensions in speech discrimination. In F. Wrestle, R.M. Shiffrin, J.J. Castellan, H. Landman, and D.B. Pisoni (Eds.), Cognitive Theory . Potomac: Eribaum.