THE ROLE OF TRANSITION VELOCITY IN THE PERCEPTION OF V_1V_2 COMPLEXES

Pierre Divenyi
Speech and Hearing Research, V.A. Medical Center, Martinez, CA, USA
Björn Lindblom
Department of Linguistics, Stockholm University, Stockholm, Sweden
René Carré
C.N.R.S. and Ecole Nationale de Télécommunications, Paris, France

ABSTRACT

V_1V_2 tokens were digitally synthesized, with the transition plus the V_2 segments cut back to various degrees. Listeners were asked to either identify the vowel at the end of the stimulus (Exp. 1) or to judge its proximity to a designated target (Exp. 2). Results suggest that the overshoot effect in the dynamic perception of vowels may be at least partially attributed to cochlear processes. Thus, while results are consistent with a gesture-bound theory of speech perception, they also support alternative accounts.

INTRODUCTION

It has been known for some time that, for a vowel embedded in a C_1VC_1 or $V_1V_2V_1$ context to be recognized, its formant frequencies do not need to reach the values characteristic to those of the same vowel in isolation [1]. This phenomenon has been termed "vowel reduction" or "perceptual overshoot." Recent results in this area overwhelmingly suggest that vowel reduction is an integral part of the production and perception of connected discourse [2][3], and it represents an accessible point of entry into the study of the dynamic aspects of speech. One aspect of vowel reduction and the associated perceptual overshoot that has not received sufficient attention is its dependence on the velocity of vocalic transitions [4]. The experiments presented in this paper investigated the effect of transition velocity on vowel perception and addressed two specific questions: (1) Can formant transitions leading to a certain vowel target define the target as a function of the transition velocity alone? (2) Is the trajectory leading from vowel V_1 to vowel V_2 perceived identically to the trajectory leading from V_2 to V_1?

EXPERIMENT 1

V_1V_2 samples were generated digitally using a PC computer. Each sample had a falling f_0-contour corresponding to that of a male voice. In Experiment 1, the two vowels were selected at either endpoint or midway between the linear trajectory between the two French vowels /a/ and /i/, represented in the F1-F2 plane. The vowel exactly bisecting this trajectory was one which French-speaking listeners identified as an acceptable token of /ɛ/. The trajectory in the F1-F2 vowel space is illustrated in Fig. 1a. The duration of the V_1 segment was held constant at 100 ms. Two velocities of frequency change were synthesized, one covering the distance between /a/ and /i/ in 100-ms, and the other in 200 ms. The trajectory thus reached the vowel /ɛ/ in 50 or 100 ms, respectively. A 100-ms terminal steady-state portion of V_2 was appended to the transition. From each V_1-V_2 complex, a series of tokens were generated by cutting back an increasingly longer segment from the end of the complex. The duration difference between any two adjacent tokens was 5 ms. Four V_1V_2 pairs, /ai/, /ia/, /aɛ/, and /iɛ/, were investigated. A schematic of a

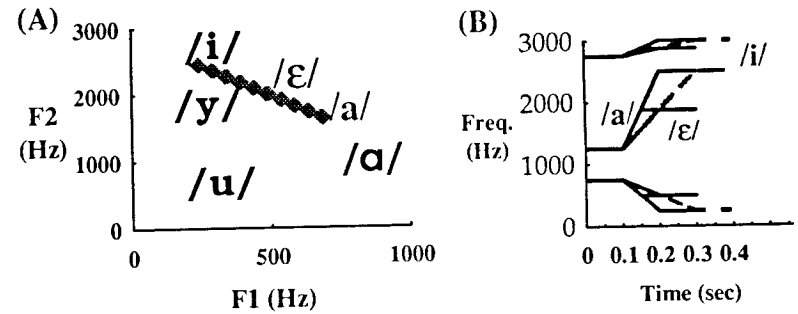

Figure 1. Stimuli used in Experiment 1. A: F1-F2 plane representation of the /il-/al transition trajectory. B: Spectrographic representation of /ail and /aɛ / at the two transition velocities (shown as different line styles).

sample stimulus is shown in Figure 1b. Three experienced subjects served as listeners. They had to indicate, by key press, which of the four vowels /a/, /e/, /ɛ/, or /i/ sounded most similar to the final vowel of the stimulus they just heard. Blocks of stimuli contained 10 repetitions of 10 to 16 different tokens. Each token was introduced by the monosyllabic word "dis" (="say").

Figures 2a-d illustrate combined results of three subjects. Each panel repre-

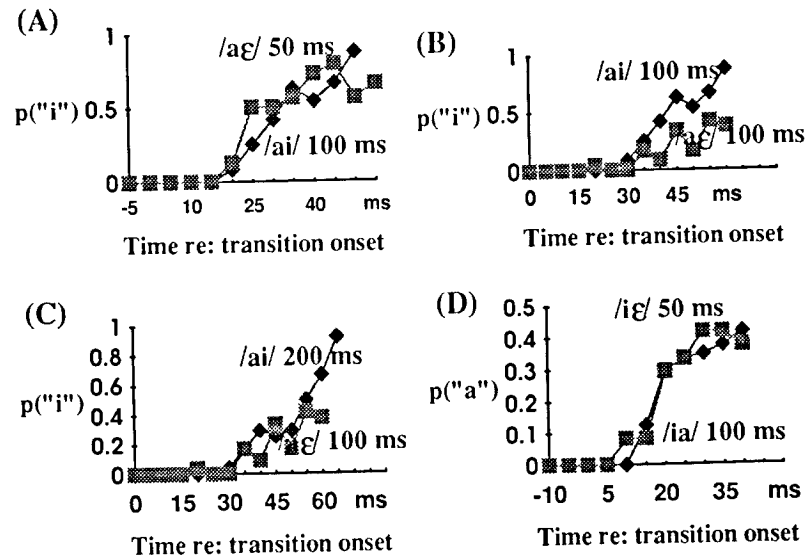

Figure 2. Results of Experiment 1. A: Percentages of a final "i" reported for /ai/ with a 100-ms transition and for /aɛ/ with a 50-ms transition (solid lines in Fig. 1b). B: Same as "A", but with 100-ms /aɛ/ transition. C: Same as "B", but with 200-ms /ai/ transition. D: Percentage of "a" reported for /ia/ with 100-ms and /iɛ/ with 50-ms transitions. Note the compressed ordinate scale. Averaged data for three listeners.

sents the proportion of the responses in which the terminal vowel segment was heard as the *remote* anchor of the trajectory, i.e., /i/ for the /ai/ and /aɛ/ transitions, and /a/ for the /ia/ and /iɛ/ transitions. The three main results are: (1) For identical Hz/ms transition rates, responses to the /ai/-/aɛ/ and /ia/-/iɛ/ pairs overlap even beyond the onset of the steady-state /ɛ/. (2) For the low

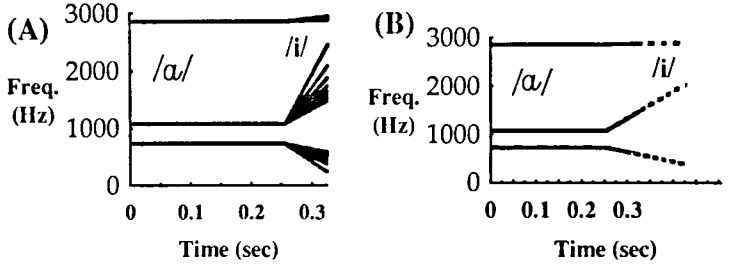

Figure 3. Spectrographic representations of stimuli used in Experiment 2. A: F1-F2-F3 of the eight /ia/ tokens used in the constant transition duration series. B: F1-F2-F3 of two of the eight /ai/ tokens used in the constant transition velocity series.

transition rate, there is little or no overshoot. (3) The /ia/-/iɛ/ and /ai/-/aɛ/ transitions are judged asymmetrically.

EXPERIMENT 2

Experiment 2 was designed to test for overshoot along trajectories outlined by the three corners of the vowel triangle with series of tokens differing *either* in transition velocity *or* in transition duration. Two series of eight tokens were generated from the six V_1V_2 complexes */ai/-/ia/*, */au/-/ua/*, and */ui/-/iu/* with the final formant frequencies of the tokens covering the V_1-V_2 formant space, with transition *duration* (Fig. 3a) or transition *velocity* (Fig. 3b) fixed. Four trained subjects served as listeners and *rated*, on a four-point scale, the *proximity* between the V_2 target and the terminal vowel segment of the token.

From the results, we estimated F1 and F2 values of the final vowel which was judged with a 50 percent confidence to be the target. Figure 4a shows these frequencies for the constant-duration and Fig. 4b for the constant-velocity series and demonstrate consistent overshoot. The largest overshoots are found in the constant duration conditions, indicating that, as the transition velocity increases, the final vowel is increasingly heard as the V_2 target. Although the overshoot for the constant-velocity vowel pairs is more modest, the variability is also less. The small size of these overshoots may be due to the averaging process that computes the pitch of sounds with changing frequencies [5]. Large directional asymmetry was observed only for one vowel pair and only in one series.

DISCUSSION AND CONCLUSIONS

The above two experiments demonstrate that, although contrast and linguistic context may influence vowel recognition, they do not represent *sine qua non* conditions for the phenomenon of overshoot to occur. The present results suggests that peripheral auditory processing may play a substantial role in the dynamic perception of vowels. In Fig. 4, it seems that the overshoot is associated with either a large F1 difference (*/ua/-/au/*) or an F2 trajectory that traverses low-frequency (<1000 Hz) regions. Since adjacent harmonics are individually resolved at low frequencies, both inter-

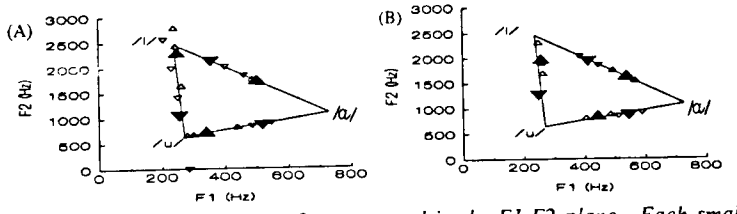

Figure 4. Results of Experiment 2 represented in the F1-F2 plane. Each small open triangle represents data from one of the four subjects and the large filled triangles, the average data for any condition. Data represent estimated formant frequencies for terminal vowels judged to reach the target V_2 with a 50-percent confidence rating. Each triangle points toward the target V_2 frequency, e.g., on the trajectory between /u/ and /i/ triangles pointing downward are data for /iu/, and those pointing upward are data for /ui/ conditions. Panel A: results of the constant transition duration series. Panel B: results for the constant transition velocity series.

harmonic and broad (especially low-by-high-frequency) *suppression* may shift the *effective* formant peak (i. e., the peak in the peripheral excitation pattern) away from another, relatively high-energy region [6].

Fig. 2 also suggests that the presence and extent of the overshoot depend heavily on the *slope* of the transition. In fact, the intended target may not even matter: An /aɛ/ doublet will generate the percept of a terminal /i/ as long as the transition velocity is identical to that of an /ai/ doublet. This illusion is strongest at the very beginning of the transition and remains quite compelling as long as a final steady-state /ɛ/ is absent. Since in the natural production of such vowel pairs the first part of the transition coincides with a period of maximum acceleration of the articulators (i.e., a period of maximum force), our data are consistent with a speech-gesture interpretation but also with certain alternative accounts.

ACKNOWLEDGEMENT

The authors would like to thank Drs. Steven Greenberg, Shinji Maeda, and John Ohala for their patiently expressed opinions on the ideas discussed in this paper. The research was supported by NIH and VA Medical Research in the U.S. and by the E.U. Science Project.

REFERENCES

[1] Lindblom, B. and M. Studdert-Kennedy (1967), "On the role of formant transitions in vowel recognition", *Journal of the Acoustical Society of America*, 42: pp. 830-843.

[2] van Son, R.J.J.H. (1993), "Vowel perception: A closer look at the literature", *Proceedings of the Institute of Phonetic Sciences, University of Amsterdam*, 17: pp. 33-64.

[3] van Wieringen, A. (1995), *Perceiving dynamic speechlike sounds*. University of Amsterdam (the Netherlands):

[4] Pols, L.C.W. and R.J.J.H. van Son (1993), "Acoustics and perception of dynamic vowel segments", *Speech Communication*, 13: pp. 135-147.

[5] Nabelek, I.V., A.K. Nabelek, and I.J. Hirsh (1973), "Pitch of sound bursts with continuous or discontinuous change of frequency", *Journal of the Acoustical Society of America*, 53: pp. 1305-1315.

[6] Weber, D.L. and D.M. Green (1978), "Temporal factors and suppression effects in backward and forward masking," *Journal of the Acoustical Society of America*, 64: pp. 1392-1399.