The Effect of Transition Velocity and Transition Duration on Vowel Reduction in V1V2 Complexes

Pierre L. Divenyi* and René Carré†

*Experimental Audiology Research, VA Medical Center, Martinez, CA 94553 and †ENST, Département Signal, 46, rue Barrault, 75634 Paris Cedex 13, France

Abstract: Transitions from a 100-ms vowel V1 toward vowel V2 were synthesized for trajectories between American English /ai/, /əi/, and /ə/ vowels. In experiment 1, the duration of the transition was constant and the velocity of the transition was different in each of eight tokens. In experiment 2, the velocity of the transition was constant and the transition duration varied from one token to the next. Subjects judged their confidence that the transition actually reached a specified target vowel. Results showed a consistent presence of vowel reduction and a trade-off between transition duration and velocity.

INTRODUCTION

Although, in connected discourse, formant transitions to and from vowels often do not reach their intended target, the listener perceives the target as if it had been, in fact, reached. This phenomenon, termed "vowel reduction" (VR) and its perceptual counterpart "perceptual overshoot" (Lindblom, 1963; Lindblom & Studdert-Kennedy, 1967), represents a readily observable instance of speech dynamics. Recent investigations of both production and perceptual aspects of VR have identified numerous characteristics of the phenomenon in diverse phonetic contexts and at different speaking rates (van Son, 1993; van Wieringen, 1995).

While VR in context can be understood as being generated by a linguistic process that strives to achieve perceptual invariance across phonetic variations resulting from variations of the context itself, existence of a parallel auditory process leading to an identical outcome cannot be altogether discounted. Indeed, there are indications that dynamic variations of the spectrum can lead to perceptual overshoot-like effects even for nonlinguistic stimuli (Divenyi, Carré, & Algazi, 1997). Thus, studying, within a limited linguistic framework, the effect of parametric changes of a transition on the perception of an unreached target could reveal hitherto hidden properties of the perceptual dynamics of both speech and hearing. The objective of the present experiments was to investigate temporal aspects of vocalic transitions, namely, the relationship between their duration and their velocity.

THE EXPERIMENTS

Stimuli were cascade-parallel formant synthesized vowels with a 100-ms V1 first portion followed by a transition following a linear trajectory in the F1-F2 plane, toward target vowel V2. Six American English V1V2-transition complexes were studied: /ai/, /əi/, /əai/, /əai/, /əi/, and /ai/. Eight tokens were generated for each vowel complex. In experiment 1, the tokens had a constant duration but the duration of the transition toward V2 was gradually increased (i.e., its velocity decreased) with the result that the final F1-F2 frequencies of the tokens approached those of V2 as the transition became shorter (see Figure 1a). In experiment 2, the transition portion of the tokens had a constant velocity but the duration of each token was gradually cut back, with the result that the final F1-F2 frequencies of the tokens approached those of the target V2 as the transition became longer (see Figure 1b). Each token was presented ten times per block of trials, one vowel complex per block. The listeners' task was to rate their

FIGURE 1. Time-frequency plots of the first two formants of the stimuli for the two experiments. Only four of the eight tokens are shown and only for the /ai/ V1V2 transitions. The vertical lines signify cutoff points.
confidence level (on a four-point scale) whether the end of the transition had reached the V2 target (presented prior to each block). An estimate of the extent of VR was computed from identification estimates derived from the rating ROC curve of each token. These VR estimates were converted to the corresponding velocity (exp. 1) or duration (exp. 2) of the transition, as shown in Figure 2. Note that an almost perfect trade-off is displayed between velocity and duration, but the relationship is nonlinear (because the slopes are not equal to –1). The only exception is the /au/ transition that was insensitive to duration (in exp. 1) and produced the smallest VR. Note also that a velocity of 0.02 oct/ms is roughly equivalent to one critical band per 10 ms, i.e., a generally accepted estimate for the peripheral auditory time window.

FIGURE 2. Average data for experiment 1 (top panels) and experiment 2 (bottom panels). F1 transition data are shown for all V1V2 sequences, except /ui/ and /ia/ for which F2 transition data are displayed. Larger velocities and shorter durations are associated with a larger degree of vowel reduction.

CONCLUSION

VR embodies a phenomenon that catches both linguistically controlled phonetic analysis and peripherally originating auditory analysis in motion. Since VR appears to represent a testable function, further investigations of this phenomenon could allow us to study both phonetic and auditory dynamics. [Supported by the VA Medical Research, NIH, and the NATO Office of Scientific Research].

REFERENCES