PROCEDURAL INFLUENCES ON THREE MEASURES OF ARTICULATORY CONTROL

Ruth Huntley, Allen Montgomery, Theresa Herron,
Heather Clark, Kathy Kugel and Laura Iden
University of South Carolina, Columbia, SC USA

ABSTRACT

This project was designed to study the stability of three measures of articulatory control: a relative timing task, the F2 slope index and locus equations. College-age subjects were asked to read traditional stimulus types at conversational, slow and fast rates at two separate testing sessions. Statistical analyses indicated that a change in rate significantly affected two out of the three measures while the test-retest factor was not significant. However, normalization procedures revealed that rate continued to be a factor in the F2 slope index while exerting little to no influence on the other techniques.

INTRODUCTION

The articulatory process is indeed a complex one that is mediated by both central and peripheral factors. Numerous investigators have attempted to identify the essential components of this process in an attempt to draw conclusions regarding the planning, programming and execution of phonemes in a sequence. Three analysis procedures have appeared in the recent literature on articulatory control that merit further examination because they offer some support to the notion of relational invariance. These techniques include a relative timing task [1,2], the F2 slope index [3,4] and the locus equation [5,6].

Relative timing refers to the inherent temporal organization of phonetic units across the time dimension. Several studies have indicated that relative timing does remain constant throughout changes in speech rate and stress patterns [1,7]. These investigators have concluded that timing characteristics are inherent in the motor programs for speech and not superimposed on the signal at the level of the articulators. The F2 slope index, on the other hand, has been used to demonstrate the importance of formant trajectories in the determination of speech intelligibility [3,8]. It is postulated that a flatter slope than "normal" would be perceived as speech that was less intelligible [4]. Finally, locus equations are used to demonstrate relational invariance (i.e., independent of vowel context) of place of articulation for initial voiced stop consonants [5]. These equations are so robust that they even hold across languages [6].

While all of these measures seem to be well established, there is little information regarding their effectiveness in conditions where the rate of speech is either faster or slower than normal. Since rate has been shown to influence segment duration and formant frequency values [9-11], it seems important to test this factor more carefully. Furthermore, little is known about the stability of such measures across times of testing. Therefore, it is the purpose of this project to assess the influences of rate of speech and test-retest reliability on each of these measures of articulatory control. Finally, as a test of intersubject variability, the investigators normalized the rate and formant values obtained in their measurement procedures. The normalization process should highlight the factors that may be changing as a function of speech rate. Since timing ratios are presumed to reflect the central organizing activity of the articulatory process, they should not be affected by changes in rate. On the other hand, the F2 measures represent movement of the articulators. Therefore, statistical normalization of rate may affect the F2 slope measures while F2 normalization (adjusting for vocal tract length) may influence F2 slope and F2 regression lines.

SUBJECTS

Twenty subjects (10 males and 10 females) participated in this study. They ranged in age from 22-33 years of age and were enrolled at the University of South Carolina. All of the subjects spoke Standard American English. They had no known speech, voice, fluency or hearing disorders and were screened for minimal dialectal variation. None of the speakers had received professional voice training; however, the talkers were given a chance to practice the tasks prior to commencement of the procedure.

PROCEDURE

The voice recordings were made with a Digital AudioTape (DAT) recorder in a sound-treated booth. The subjects were presented with a set of cards containing the stimulus items. Each of the speakers read the cards aloud into the microphone. They were required to produce each stimulus set, starting with the first phrase and continuing on to the last one, before moving on with the next repetition. The subjects read the stimulus sets first at a normal, conversational rate, followed by a slow rate and finally at a fast rate. To achieve a "slow" rate, subjects were asked to speak at what they considered to be half their normal rate of speaking. They were also provided with a spoken model by the experimenter, as well as cues to slow down throughout the procedure. Similarly, to achieve a fast rate, subjects were asked to speak at twice their normal rate, with a model again provided by the experimenter. The entire experimental procedure was repeated two days later as a measure of test-retest reliability.

STIMULUS SETS

They were asked to read the following stimulus types: three repetitions of five sentences constructed for the ratio task [1], three repetitions of a standard phrase with a target word taken from the acoustic signature literature [3] and five repetitions of /b,d,g/ paired with four vowels in a /CVt/ context [5].

DATA ANALYSIS

All measurements were made from computer-generated spectrograms using Sensimetrics SpeechStation (version 3.0) software on an IBM-compatible 486 computer. The Sensimetrics program allows the user to identify various acoustic correlates, such as time (in msecs.) and frequency corresponding to specific points on the stimuli.

Measurement Procedures

For the relative timing ratios, the investigator constructed a set of nine measurements (one durational and four sets of ratio measurements) by which each utterance was analyzed. The ratio measurements were based on constructs developed by Weismer, et al. [1] and Prosek, et al. [2]. The boundaries for each ratio occurred at a vowel-consonant (VC) or consonant-vowel (CV) interface.

The F2 slopes were derived by identifying the last frequency value of F2 before the transition and the value of the first glottal pulse of the leveling off point of the steady state portion. The procedure was somewhat different for "wax" and "blend" in that the glides were also measured as they rose into the vowels. This procedure seemed to include the most frequency change. Once the starting and stopping points were identified, the frequency and millisecond values for each of these points were recorded and a slope was computed by dividing the amount of frequency change by the duration of that change (i.e., rise over run).

The locus equations were generated as follows: measurements were taken at the first glottal impulse of F2 and then during the vowel steady state. Each stimulus item was measured three times at each rate and joined on a scatterplot which represented each consonant by plotting the F2 onset by the F2 steady state frequencies of all four vowels. The regression line, which results from this procedure, is known as the locus equation. From these scatterplots, slopes were derived for intersubject comparisons. Lastly, all measurements were subjected to intra- and inter-judge reliability testing and all correlations derived from these procedures were excellent (r > 0 .85).

Normalization Procedures

Rate normalization was carried out on the timing ratios and the F2 slope measurements. The timing ratios were recomputed for the fast and slow rates with corresponding durations from the normal rate serving as the denominator. These new component fractions were used to calculate "normalized" ratios. The intent was to factor out the influence of rate across the ratio portions. If speech rate was a constant factor across the generation of a phrase, the ratios generated for each component should be equal and the ratio should be equal to one. Thus, normalized ratios that deviated from one would indicate differential increase (or decrease) in rate across the sentence.

In adjusting the rate of F2 slopes, the investigator utilized the "normal" duration value as the denominator for all of the slopes. This process would provide information about F2 rise while holding time constant. Now, one could talk about the influences of coarticulation independent of rate.

F2 frequency normalization was also employed as a means of controlling for individual differences in estimated vocal tract length. Therefore, the extent of frequency change noted here could provide insight into the process of coarticulation while controlling for one factor known to vary across individuals.

RESULTS AND DISCUSSION

Two or three-way ANOVAs (*stimuli x* rate x testing time) were conducted on the preliminary data to determine the effects of rate and test-retest reliability on each of these measures of articulatory control. Rate was noted to be a significant factor in two out of three of these measures. However, the test-retest factor was not significant. Therefore, one can conclude that the articulatory process associated with each of these measurements is fairly consistent across times of testing.

Effects of Speaker Rate

The use of a slow rate proved to be troublesome for the relative timing ratios in that three of the four ratios did not remain constant in this condition. Furthermore, one of the sentences generated variable ratios in both the fast and slow conditions for at least two of the ratios. While the sentences were chosen to represent different types of phoneme transitions, this factor did not seem to be the most pertinent. However, the semantic naturalness of the sentence seemed to disrupt relative timing. This factor needs further investigation.

Rate, once again, was a factor in the interpretation of the F2 slopes. The mean slopes for each rate were significantly different from one another. Only the analysis of the words "wax" and "blend" provided different results. In these cases, the mean slope values for the normal and fast rates were not found to be significantly different from one another; however, slope values for the slow rate were distinctive from the other two speeds.

Finally, a change in rate did not affect the derivation of a locus equation for any of the voiced plosives tested. This finding was to be expected since the locus equation is not a time dependent measurement. However, it is interesting to note that the possible target undershoot and overshoot that occurred in the fast and/or slow conditions did not significantly affect the slope of the regression line for /b/, /d/ or /g/. Furthermore, as noted in previous research, the slope of the /d/ regression line was significantly different from the others.

Effects of Normalization Procedures

Rate normalization revealed that the effects of rate on the relative timing ratios are consistent throughout the sentence with mean ratios for both fast and slow conditions approximating 1.0. That is, when differences attributable to talker variation and sentence context were removed, the ratios revealed no differential lengthening or shortening of sentence components on the average.

However, specific sentences or ratios showed some modest effects of rate.

As might be expected, rate normalization of the F2 slope reduced differences attributable to talker speed. However, significant differences for the fast and slow rates for certain words still remained indicating the continuing presence of articulatory over/undershoot as a contributory factor.

The F2 frequency normalization procedures are currently underway and no conclusions are available at this time. In general, the three measures employed in the present study appear to be useful tools for assessing articulatory behavior and further refinements and enhancements of these techniques would appear to be justified.

REFERENCES

[1] Weismer, G. & Fennell, A. (1985), "Constancy of (acoustic) relative timing measures in phrase-level utterances," *J. Acoust. Soc. Amer.*, vol. 78, pp. 49-57.
[2] Prosek, R., Montgomery, A. & Walden, B. (1988), "Constancy of relative timing for stutterers and nonstutterers," *J. Sp. Hear. Res.*, vol. 31, pp. 644-658.
[3] Kent, R., Weismer, G., Kent, J. & Rosenbek, J. (1989), "Toward phonetic intelligibility testing in dysarthria," *J. Sp. Hear. Dis.*, vol. 54, pp. 482-499.
[4] Kent, R., Kent, J., Weismer, G., Martin, R., Sufit, R., Brooks, B. and Rosenbek, J. (1989), "Relationships between speech intelligibility and the slope of second-formant transitions in dysarthric subjects," *Clin. Ling. and Phon.*, vol. 3, pp. 347-358.
[5] Sussman, H., McCaffrey, H. & Matthews, S. (1991), "An investigation of locus equations as a source of relational invariance for stop place categorization," *J. Acoust. Soc. Amer.*, vol. 90, pp. 1309-1325.
[6] Sussman, H., Hoemeke, K. & Ahmed, F. (1993), "A cross-linguistic investigation of locus equations as a phonetic descriptor for place of articulation," *J. Acoust. Soc. Amer.*, vol. 94, pp. 1256-1268.
[7] Tuller, B., Kelso, J. & Harris, K. (1983), "Converging evidence for the role of relative timing in speech," *J. Acoust. Soc. Amer.*, vol. 76, pp. 1030-1036.
[8] Weismer, G., Kent, R., Hodge, M. and Martin, R. (1988), "The acoustic signature for intelligibility test words," *J. Acoust. Soc. Amer.*, vol. 84, pp. 1281-1291.
[9] Lindblom, B. (1963), "Spectrographic study of vowel reduction," *J. Acoust. Soc. Amer.*, vol. 35, pp. 1773-1781.
[10] Gay, T. (1968), "Effects of speaking rate on diphthong formant movements," *J. Acoust. Soc. Amer.*, vol. 44, pp. 1570-1573.
[11] Gay, T. (1978), "Effect of speaking rate on vowel formant movements," *J. Acoust. Soc. Amer.*, vol. 63, pp. 223-230.