THE DISTINCTION OF CENTRAL AND PERIPHERAL SPEECH TIMING MECHANISMS

Eric Keller

Département de linguistique, Université du Québec à Montréal, Montréal, Qc. H3C 3P8, Canada

ABSTRACT

This study examines the multiple and conjoint prediction of speech timing events by central and more peripheral mechanisms. Phonemic ("central") distinctions showed greater predictive power for VOT segments, while rate ("more peripheral") distinctions showed greater predictive power for syllable intervals and vocalic durations. In patients with cerebellar disorders ("ataxic dysarthrics", patients suffering from a "relatively peripheral" motor disorder), the predictive power of speech rate was more strongly impaired than that of consonant distinction.

1. INTRODUCTION

Traditionally, phonetic science has considered variability to be a nuisance variable. This has been particularly so with respect to timing, where considerable variability is observed in intra- and inter-syllabic durations over repeated productions of the same utterance by the same or by different speakers.

However in line with most contemporary behavioral and social sciences, an alternative theoretical approach to variability is possible. In this view, variability is the result of the combined effects of a multiplicity of factors, some of which may be related to central speech processing, others to more peripheral motor processing and yet others to muscular execution (Figure 1).

In the domain of speech timing, a variety of potential predictors can be proposed for time segments measured at the periphery (e.g., in an acoustic waveform). On the one hand, lengthening or shortening effects can be due to linguistic factors, such as semantic emphasis, syntactic pauses, or phonemic distinctions. On the other hand, some timing effects are related to overall speech rate and to rhythmic variations.

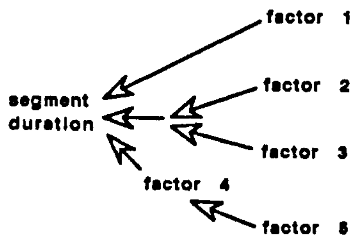

Figure 1. A theoretical approach to explaining variability in speech timing. Variability is seen as the outcome of the combined effects of a multiplicity of factors.

The present experiment examines the predictive interplay of two such factors. The first factor is phonemic distinctiveness (specifically, the ternary distinction between /p/, /t/, and /k/ in CV syllables). Since this distinction is relevant to lexemic distinction, it is considered to be representative of *linguistic control*.

The second factor is speech rate (e.g., normal or fast rate in a simple, repeated CV paradigm like /papapa/). Since many repeated motor actions such as walking and tapping can be produced at a faster or slower rate, this factor is considered to be representative of *general motor control*.

Normal speech probably involves concurrent processing at linguistic and general motor control levels. Therefore,

the two types of factors should exert a combined influence on durational segments in speech. In addition, the linguistic control factors should predict the greatest proportion of variance in those time segments that serve most directly in the acoustic distinction of syllables, such as VOTs. Conversely, general motor control factors should predict the greatest proportion of variance in other time segments in speech.

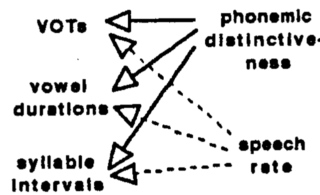

Figure 2. The predictive pattern examined in this study. Three phoneme categories (/p/, /t/, /k/ in CV syllables) and two speech rates (normal, fast) predict the duration of three speech periods (VOTs, vowel durations, syllable intervals).

A further test of this theoretical framework is possible. Patients suffering from neurological lesions affecting predominantly general motor control should show the greatest reduction in predictive power in speech rate. After all, if speech rate is indeed processed by a structure similar to that which controls the rate of production of other motor actions, an impairment affecting such a structure should have similar effects on speech motor control as on limb control.

2. METHOD

Seven dysarthric patients with cerebellar and/or ponto-cerebellar lesions (mostly diagnosed as Friedreich's ataxia) and six control subjects were asked to produce either /pa/, /ta/ or /ka/ stimuli repeatedly until the examiner held up his hand (minimum: 5 seconds). Tasks were performed at fast and conversational speech rates. Patients had been selected from a larger group of 13 patients for their particular severity of impairment.

Recordings were digitized at 10.4 kHz with a 12-bit MacAdios Model 411 system. Time measures were taken at three points in the acoustic waveform (see Figure 3), and three speech segments were calculated from these measures. Points 1 and 2 are defined in traditional manner for VOT at the burst and at the onset of voicing. Point 3 is defined by the loss of vocalic oscillation, as judged against a noise threshold of the succeeding resting signal segment. Three representative durational measures derived from these observation points (VOT, vowel duration and syllable interval) were selected from an original 10 time measures.

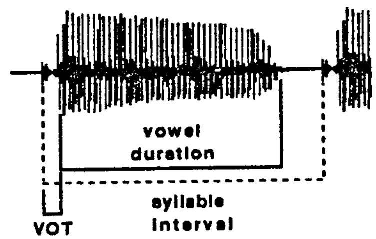

Figure 3. Duration measurements on the acoustic waveform. Out of ten durational measures performed within syllables and between adjoining syllables, VOTs, vowel durations and syllable intervals were retained as representative time segments.

In the subset of the data discussed here, there were 5,733 observations (out of an original 23,586 measurement points). Interjudgemental agreement on 2,880 remeasured pairs of measurement points was 98.6%.

Because of moderate to severe positive skewness, all measured time segments were log transformed, then z transformed, and measures exceeding ±3.0 s.d. were eliminated (33 out of 23,586, or 0.13%). Subsequently the probability that data was not normally distributed was $< .05$.

Standard multiple regressions of the form:

speech segment =
phoneme category
+ speech rate category
+ constant

were performed separately for each subject or patient, as well as for each measured speech segment (39 cells). There were an average of 147 observations per cell. Predictors were ternary-valued for the phoneme category [/p/, /t/, /k/] and binary-valued for the speech rate category [normal, fast]. Degree of prediction was derived from the multiple regressions' averaged absolute beta coefficients for each type of predictive relationship.

3. RESULTS

The results of the regression analyses were in agreement with the hypotheses specified above.

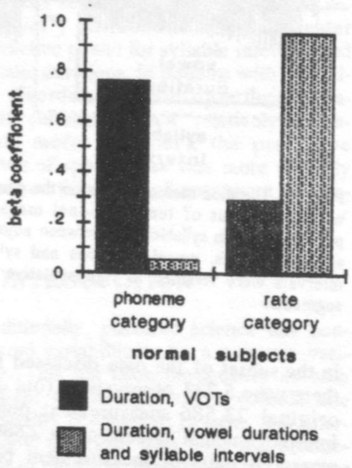

Figure 4. The prediction of time segments in six normal subjects. On the average, the ternary phoneme distinction showed the best prediction for VOT durations, while the binary speech rate distinction had the best prediction for vowel durations and syllable intervals. Results also indicate that time segments tended to be *jointly* predicted by phoneme and rate differentiations (illustrated in this figure for the prediction of VOT). Results for vowel durations and syllable intervals were similar throughout the study and were combined for presentation here.

(1) *The hypothesis of joint prediction.* Results for the normal subjects showed that time segments tended to be *jointly* predicted by phoneme category and speech rate category (Figure 4). Although the predictive relations were weak in some cases (e.g., a beta of 0.057 for the relation "phoneme category predicts vowel duration/syllable interval"), the majority of the 400 original cells (77.3%) showed predictive relations significant at $p<.05$, and most (63.5%) were significant at $p<.001$, indicating that the two predictors tended to co-vary with all of the three time measures.

(2) *The hypothesis of distinct prediction.* The phoneme category had good predictive power for VOTs (beta 0.762), while rate category had excellent explanatory power for vowel durations and syllable intervals (beta 0.955, Figure 4). Crossed correlations ("speech rate predicts VOT" [beta 0.294], and "phoneme distinction predicts vowel durations and syllable intervals" [beta 0.057]) showed less predictive capacity.

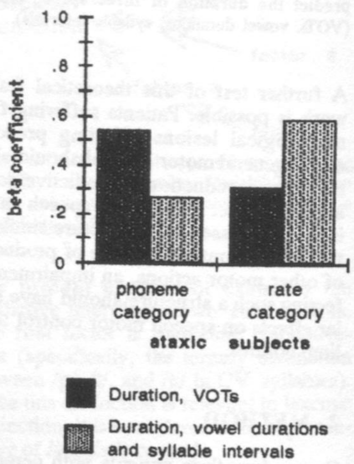

Figure 5. The prediction of time segments in seven patients with cerebellar disorders (patients with ataxic dysarthria). In comparison to the normal subjects, the (more peripheral) cerebellar disorder affected the predictive capacity of speech rate more strongly than the predictive capacity of consonant distinction. This offers some support for the notion that phonemic distinctions are part of a central programming mechanism, while speech rate is more directly related to a general motor programming mechanism.

(3) *The hypothesis of select impairment.* The prediction for patients with impairment of general motor control is also supported. It was expected that such patients would show a greater reduction of control over the relation "speech rate predicts vowel durations and syllable intervals", and a lesser impairment of the relation "phoneme category predicts VOT". Indeed, the first type of prediction was more diminished than the second (a reduction of 39%, beta 0.585 instead of beta 0.955, Figure 5). The relation "consonant distinction predicts VOT" showed a reduction of 29% (beta 0.541 instead of beta 0.762).

4. DISCUSSION

The present experiment illustrates a small fraction of the entire framework of predictive relations that is likely to characterize timing relations in speech.

The results support a view that considers phonetic variability at the periphery to be the predictive outcome of a multiplicity of factors, including linguistically relevant determiners like phonemic distinctiveness and general motor control determiners like speech rate. There is also some support for the notion of considering some of these factors, like the linguistic factors, to be more "central" and others, like the general motor control factors, to be more "peripheral" in nature.

The view supported by the present data thus contradicts earlier approaches to speech timing that attempted to view timing variations due to changes in rate and stressing as simple metrical variations of a basic temporal organization (the "proportional timing" hypothesis proposed predominantly by Kelso and colleagues, e.g. [2], see also [1] and [3]).

At the same time, the interesting, but somewhat limited results (39% against 29% reduction) from the pathological populations induce some caution. It is recalled that the present group of seven patients with presumed cerebellar and ponto-cerebellar lesions had been selected for the great severity of their impairment. Although these are patients whose cortical processing should not be affected by a direct lesion, their relation "phoneme category predicts VOT" was also reduced (by 29%). And while their excessive timing variabilities and great difficulties in controlling limb movement betrays extensive cerebellar impairment, 11 of 14 cells illustrating the relation "rate category predicts vowel duration and syllable interval" were still significant at $p<.05$ (8 of 14 at $p<.001$). Lesions affecting a portion of the motor output system presumably interferes with the entire system, particularly the processing of events "upstream". At the same time, lesions presumably affecting a specific process rarely succeed in obliterating its entire functionality.

Finally, the study illustrates one of several interesting statistical techniques that can be used to explore the complete timing framework. Multiple regression and its more sophisticated outgrowth, path analysis, would seem to be the natural analysis techniques for a complex structure consisting of multiple predictor categories and a large number of predicted time measures in the speech utterance.

5. REFERENCES

[1] KELLER, E. (1990). Speech motor timing. In W.J. Hardcastle and A. Marchal, *Speech Production and Speech Modelling* (pp. 343-364). Amsterdam: Kluwer.

[2] KELSO, J.A.S., SALTZMAN, E.L., & TULLER, B. (1986). The dynamical perspective on speech production: Data and theory. *Journal of Phonetics, 14*, 29-59.

[3] MUNHALL, K.G. (1985). An examination of intra-articulatory relative timing. *Journal of the Acoustical Society of America, 78*, 1548-1553.