Qualitative acoustic analysis in the study of motor speech disorders

Julie M. Liss
Department of Communication Disorders, 115 Shevlin Hall, University of Minnesota, Minneapolis, Minnesota 55455

Gary Weismer
Department of Communicative Disorders and the Waisman Center, University of Wisconsin, Madison, Wisconsin 53705

(Received 9 March 1992; accepted for publication 29 July 1992)

Traditional measurements performed on the acoustic signals of normal speech are frequently used to quantify the acoustic characteristics of disordered speech as well. This letter demonstrates how important aspects of speech production deficits in motor speech disorders may be overlooked if stringent quantification procedures are employed, especially in the stage of exploratory data analysis. It is suggested that qualitative procedures, wherein phenomena are inferred from visual examination of certain acoustic displays, are useful to supplement traditional measurements, and moreover, that they be used to point to the types of measurements that should be made in the finer-grained stages of quantitative analysis.

PACS numbers: 43.70.Dn

INTRODUCTION

We have asserted elsewhere that traditional acoustic measures of temporal and spectral characteristics of normal speech may not necessarily reveal the inherently "important" aspects of disordered speech production (Weismer and Liss, 1991; see also Westbury, 1991). By "important," we mean those characteristics of the acoustic signal that are likely to reflect aspects of disordered sensorimotor control and/or perceptual phenomena that will play a prominent role in a theory of motor speech disorders. We have also argued that because the traditional parametric approaches necessarily exclude idiosyncratic and aberrant speech production from the analyses, and in fact may be compromised by the variability, much potentially revealing information is sacrificed (Sussman et al., 1988).

The purpose of this letter is to describe how the inclusion of a qualitative level of acoustic analysis can elucidate potentially "important" phenomena in the disordered production of vocalic segments, and how this can lead to theoretical notions that yield testable hypotheses in both the quantitative and qualitative domains. By "qualitative" level of analysis, we mean a visual inspection and interpretation of acoustic data, especially involving repetitions of utterances by individual speakers. The examples described in this letter were obtained in the context of a larger investigation of the acoustic effects of contrastive stress among normal geriatric men, and subjects exhibiting apraxia of speech and ataxic dysarthria (Liss and Weismer, submitted). Whereas the current examples involve qualitative analyses of formant trajectories associated with vocalic nuclei, the analysis approach might be applied to any set of acoustic or kinematic data (see Cooke and Brown, 1986, for a discussion of a similar method in research on limb motor control, and Corcos et al., 1990, for examples of such analyses).

I. METHODS AND PROCEDURES

For the larger study, subjects included four each of control (C), apraxic (A), and ataxic dysarthric (D) speakers. Detailed descriptions of these subjects can be found in other published works (e.g., McNeil et al., 1990). For the purposes of demonstration, we will examine data from one normal male subject (age 63), and two male speakers with apraxia of speech (age 54 and 62) who were free of significant concomitant dysarthria or aphasia.

Subjects produced phrases and sentences following tape-recorded stimuli. These productions were recorded on cassette tape using high fidelity equipment. The two utterances analyzed in this investigation, "buy Bobby a poppy," and "build a big building," were produced five times each in two conditions (for a total of 40 test utterances per subject). These were randomly distributed among other sentences and phrases. Two speaking conditions were utilized in the larger study: neutral and contrastive stress. Each sentence was produced in a neutral condition, for which no specific directions about stress placement were given, and in a condition in which each of the content words was contrastively stressed.

The quantitative acoustic analysis of the larger investigation consisted of measurement of segment and utterance durations, and measurement of formant transition characteristics. Here we will discuss only formant transition characteristics for /aI/ ("buy") and /ll/ ("build"). These segments were selected because they are associated with relatively large and complex changes in articulatory configuration that are reflected most notably in the trajectory of the second formant frequency (F2). We use the term "trajectory" to refer to the time-frequency path of the center of the formant band across the entire duration of the vocalic nucleus. Transition duration (TD, in ms) and transition extent
(TE, in Hz) of the most rapidly changing portions of the trajectories were measured using a Kay DSP 5500 workstation using a wide-band (300 Hz) spectrographic display (0–4 kHz scale expansion) and the associated waveforms (see Weismer and Liss, 1991). Slope values were calculated from these measures (TE/TD). Slope values were not calculated for productions that did not contain the expected trajectory shape [e.g., Fig. 1(b), production #2 does not contain a 20-Hz change in any 20-ms segment—our operational definition of “flat” (see Weismer et al., 1988)]. Inter- and intra-judge reliability was acceptable for all quantitative measures (Liss and Weismer, submitted).

In the qualitative phase of the analysis, multiple trajectory tracings were superimposed and visually examined to identify and describe patterns and phenomena. This technique was originally designed to accommodate formants that were excluded from quantitative assessment because they could not be measured according to our operational definitions. Such formants typically corresponded with aberrant productions of the vowel that did not yield the expected trajectories for the segments of interest. In the figures shown here, there are five trajectories plotted per panel, corresponding to five repetitions of a particular utterance/condition. Mean slope values (from the quantitative analysis) and a transcription of each of the syllables containing the plotted formant trajectories are also included in the figures. The transcriptions are included only to describe the general segmental characteristics of the syllables, and not as indices of the normality or aberrancy of the productions.

The trajectories shown in this paper (Figs. 1 and 2) are used to illustrate how the qualitative procedure can be used to identify and catalogue phenomena at the individual subject level of analysis. Our point is that this kind of qualitative examination can expose phenomena that must be explained in theoretical accounts, and in fact takes advantage of intra- and intersubject variability as objects of theoretical interest, as opposed to a view wherein that variability is an obstacle to successful statistical treatment of the data.

II. RESULTS AND DISCUSSION

A. Temporal translocation and gesture scaling

Figure 1(a) displays five F2 trajectories for the segment /ll/ (from “build”) produced by a control subject, and Fig. 1(b) shows F2 trajectories for the same segment produced by a subject with apraxia of speech.

FIG. 1. (a) Five F2 trajectories for the syllable nucleus /ll/ produced in the “BIG” stress condition by a neurologically normal subject; (b) five formant trajectories for the same syllable and stress condition produced by a subject with apraxia of speech.

FIG. 2. (a) Five F2 trajectories for the syllable nucleus /ll/ produced in the “BUILD” stress condition by a subject with apraxia of speech; (b) five formant trajectories produced by the same subject in the “BIG” stress condition.
by an apraxic speaker. All of these trajectories were taken from utterances in which the word "big" was contrastively stressed.

Using these two sets of trajectories, it is possible to demonstrate how traditional measures of formant slope and segment duration can under-represent important differences between these sets of formants. Consider that the mean slope value of the steepest portions of the trajectories from the normal speaker [Fig. 1(a)] is 9.62 Hz/ms (s.d. = 1.48). Compared to the mean slope (5.23, s.d. = 0.81) of four of the trajectories produced by the apraxic speaker in Fig. 1(b), we can conclude that the trajectories of the control speaker are steeper. Further comparison reveals that the transition durations of the apraxic subject are also generally greater than those of the normal speaker. Thus, from this traditional quantitative approach, we can infer that the major articulatory gesture for /I/ was produced at a slower rate and over a greater period of time, as compared to the gesture for the normal speaker.

This conclusion is consistent with other studies of the articulatory deficit in apraxia of speech (cf. Kent and Rosenbek, 1983), and one could be inclined to stop here. However, the conclusion of slowness and the possible lengthening of the major articulatory gesture does not capture all that is unusual about the trajectory repetitions. Despite the general similarities among the most rapidly changing portions of the trajectories, there are substantial differences among the durations of the segments that precede them. In Fig. 1(a), the segment of the trajectory preceding the downward sloping portion is relatively brief and topologically uniform across all of the five tokens. This is not true for the repetitions of the apraxic speaker [Fig. 1(b)] where the trajectories appear to be "translocated" across the time axis. Specifically, these trajectories generally end in the expected downward slope, but the durations of the flat portions preceding the downward segments range from approximately 25 to nearly 200 ms. These plateaus would be consistent with periods of relative articulatory immobility, and must be explained in any comprehensive theory of motor speech disorders. The point here is that without visual examination of these superimposed formant trajectories, one would not think a priori to measure the time preceding the downward sloping portion of the trajectories. Thus, the qualitative step of visual examination of the disordered trajectories reveals a phenomenon that lends itself easily to quantification; that is not apparent from the tracings of a normal speaker; and that may be associated with underlying issues of motor control.

Another example can be found in Fig. 2(a). Note that four of the trajectories have essentially the same starting frequency (around 1300 Hz). Nevertheless, the onsets of the downward sloping transition in this set of trajectories—the "temporal translocation" of the transitions—are highly variable, apparently as a consequence of variability in gesture scaling following release of the /b/. These gesture scaling difficulties, as evidenced here by large variations in the magnitudes of the initial rising frequency swings along the F2 trajectories, should be associated with variations in the onset time of the major articulatory gesture (i.e., the transition). In other words, temporal translocation of transitions is probably related to a gesture scaling problem that is expressed in the form of variability of the gesture magnitude. A direct quantitative prediction of this view is that hyperscaled gestures (those with large initial swings in F2) should be associated with later-occurring onsets of the major articulatory gesture. Examination of the five trajectories in Fig. 2(b) generally supports this idea. Here again, the development of such predictions is based largely on visual inspection of the superimposed trajectories, and would not emerge from strict adherence to traditional quantification.

B. Articulatory decomposition/segmentalization

The preceding examples illustrate that observations from qualitative analysis (temporal translocation) can lead to testable hypotheses about the underlying mechanism (gesture scaling). Taking this one step further, the qualitative analysis can also point to hypotheses that bear directly on the development of theories of motor speech disorders.

The scaling problem described above and its possible relationship to temporal translocation of transitions may also be related to, or be a byproduct of, articulatory decomposition (or, segmentalization) of speech. This phenomenon is thought to be characterized by a reduction in the degree of overlap between successive articulatory gestures (Weismer and Liss, 1991), resulting in speech acoustic waveforms that reflect reduced coarticulation or coproduction (i.e., evidence of motor control deficits). Articulatory decomposition has been observed previously for both apraxic and dysarthric speakers (Kent and Rosenbek, 1983; Weismer and Liss, 1991; Weismer et al., 1992), and is likely to be largely responsible for the perception of "scanning speech" (reduced segment duration contrasts) in these disorders (e.g., Ziegler and von Cramon, 1986). Compare the productions of an apraxic speaker shown in Fig. 2(a) and (b) to those of the control speaker in Fig. 1(a). Four of the five F2 values at the onset of the normal trajectories, or at the point where the consonant constriction is just released into the vocalic segment, range between 1600-1700 Hz. This is a reasonable "target" value for /I/ produced by a geriatric male in this phonetic context. Four trajectories show a brief steady state (about 30 ms) around these frequencies, followed by the characteristic falling transition. In contrast, seven of the ten /I/ trajectories produced by apraxic speaker [Fig. 2(a) and (b)] have starting frequencies that are substantially lower than expected for /I/, due to the failure to position the tongue in a high-front position during the period of vocal tract closure for the word-initial /b/. At the release of the /b/, then, the tongue must move to the required high-front position, a gesture reflected in several of these trajectories as the initial rising portion of the F2 trajectory. These various attempts result in different scalings of the gesture, as well as the phenomenon identified above as "temporal translocation" of transitions.

C. Theoretical considerations

In our experience, the kinds of phenomena described in this paper are common in the speech of individuals with motor speech disorders. We have shown how a qualitative anal-
ysis procedure (i.e., visual examination of superimposed trajectories) can lead to theoretical perspectives on the nature of the speech production deficit in motor speech disorders and yield some testable (and quantifiable) ideas. Specifically, among the articulatory aberrancies identified here (temporal translocation of transitions, inappropriate scaling of gestures, articulatory decomposition/segmentalization), we postulate that the former two are probably byproducts of the latter. One possibility is that segmentalization of articulatory gestures induces compensatory responses on the part of the speaker that lead to variations in gesture scaling, and thus temporal translocation of transitions.

This is an attractive theoretical notion for at least three reasons. First, it attempts to unify the explanatory apparatus of various spatio-temporal articulatory difficulties in some motor speech disorders by means of a general, well-attested phenomenon (i.e., segmentalization). Second, it fits in with certain contemporary accounts of successive articulatory overlap—namely, the notion of gesture sliding and blending described by Saltzman and Munhall (1989)—that might permit theoretical analysis of motor speech disorders within the framework of a theory of normal speech production. Third, as noted above, it suggests certain empirical tests that are amenable to quantitative analysis.

The full potency of qualitative analyses likely will be realized when such attempts are guided—but not limited—by theoretical considerations. In efforts to identify the potentially "important" acoustic measures in motor speech disorders, we regard the qualitative procedure as indispensable.

ACKNOWLEDGMENTS

This research was supported by a University of Minnesota Graduate School Grant-in-Aid of Research, Artistry, and Scholarship, and NIH Grant No. NS18797. We extend our gratitude to Karen Forrest for her thoughtful comments on an earlier version of this paper. We also wish to acknowledge the laboratory assistance provided by Kristin Little and Stephanie Hanson. Address reprint requests to Julie Liss, Ph.D., Department of Communication Disorders, 115 Shevlin Hall, University of Minnesota, Minneapolis, MN 55455.

1 Contrastive stress is a common therapy technique whereby a speaker is taught to emphasize the content or important words in an utterance to enhance intelligibility and message transfer.

2 By "major articulatory gesture" we mean the part of the vocalic nucleus that demands the most extensive (and perhaps the most rapid) change in vocal tract configuration to produce the desired sound sequence. The acoustic correlate of this is the operationally defined transition portion of the formant trajectory.

3 "Topologically uniform" implies a consistent trajectory shape across repetitions. The implication is not that the specific frequency-time coordinates are identical or nearly so across repetitions, but that the different trajectories could be translated up or down the frequency scale (or, in some cases, the duration scale) with the resulting superimposition showing a minimum variability for a particular portion of the trajectory. In normal speakers, the amount of frequency or time scale translation required to demonstrate topological uniformity is typically very minor. Trajectories that are not topologically uniform, by this definition, would not show a reduction of the trajectory variability when this scale translation is performed.


