2-1-2000

On the Assessment of Stability and Patterning of Speech Movements

Anne Smith
Purdue University

Michael T. Johnson
Marquette University, michael.johnson@marquette.edu

Clare McGillem
Purdue University

Lisa Goffman
Purdue University

Published version. *Journal of Speech, Language, and Hearing Research*, Vol. 43, No. 1 (February 2000): 277-286. DOI.
Michael T. Johnson was affiliated with Purdue University at the time of publication.

On the Assessment of Stability and Patterning of Speech Movements

RESEARCH NOTE

Anne Smith
Michael Johnson
Clare McGillem
Lisa Goffman
Purdue University
West Lafayette, IN

Speech requires the control of complex movements of orofacial structures to produce dynamic variations in the vocal tract transfer function. The nature of the underlying motor control processes has traditionally been investigated by employing measures of articulatory movements, including movement amplitude, velocity, and duration, at selected points in time. An alternative approach, first used in the study of limb motion, is to examine the entire movement trajectory over time. A new approach to speech movement trajectory analysis was introduced in earlier work from this laboratory. In this method, trajectories from multiple movement sequences are time- and amplitude-normalized, and the STI (spatiotemporal index) is computed to capture the degree of convergence of a set of trajectories onto a single, underlying movement template. This research note describes the rationale for this analysis and provides a detailed description of the signal processing involved. Alternative interpolation procedures for time-normalization of kinematic data are also considered.

KEY WORDS: speech production, motor control, movement analysis

A widely used strategy for gaining insights into the control of movement is to search for invariance in movement trajectories. This approach has been used in the study of many different movement production systems, from octopus tentacle movements to human arm and speech movements (Atkeson & Hollerbach, 1985; Flash & Hogan, 1985; Gutfreund et al., 1996; Ostry, Cooke, & Munhall, 1987; Smith, Goffman, Zelaznik, Ying, & McGillem, 1995). When kinematic invariance is sought, two related aspects of movement trajectories are often assessed: (1) the degree to which a set of trajectories shows stereotypic features (e.g., a bell-shaped velocity profile), and (2) how variable a set of trajectories is in relation to this standard pattern.

The majority of studies of speech kinematic output have employed measures at single time points (e.g., Ackermann, Hertrich, & Scharf, 1995; Kent & Moll, 1975; Kuehn & Moll, 1976; Zimmermann 1980a, 1980b) to search for invariant aspects of motor output. In these studies, rather than considering the movement trajectory as a whole, specific points are selected to characterize temporal and spatial aspects of motion. In a smaller number of studies, movement trajectories for single speech movements were analyzed to determine if there is a common pattern in the velocity profile (Adams, Weismer, & Kent, 1993; Ostry et al., 1987; Shaiman, Adams, & Kimelman, 1997). Thus earlier work focused on single points in time to represent fundamental kinematic

parameters of movement (e.g., displacement, peak velocity, and duration), and a few investigations attempted to determine if the bell-shaped velocity profile, prevalent in many limb movements, also characterized single speech movements. In 1995, we introduced an analysis that employed the entire lower lip movement trajectory for a six-syllable phrase (Smith et al., 1995). After linearly amplitude- and time-normalizing each multicomponent movement trajectory, an average trajectory for the set of trials within one condition was computed. Standard deviations of the set were computed as a function of normalized time. The average trajectory reveals aspects of the underlying *pattern* of movement, whereas the cumulative sum of the standard deviations (the spatiotemporal index, STI) indicates the degree to which the set of trajectories converges on a single underlying template, or the *stability* of the movement sequences.

Since publication of that initial work, we have used these analytic techniques to examine a number of issues in speech motor control, including changes in patterning and stability related to (a) alteration of a single phoneme (Goffman & Smith, 1999), (b) maturation over the childhood years (Smith & Goffman, 1998; Goffman & Smith, 1999) and aging (Wohlert & Smith, 1998), (c) increased linguistic processing demands (Maner, Smith, & Grayson, in press), and (d) stuttering (Kleinow & Smith, in press). We have found the technique of linear normalization followed by computation of a composite index of spatiotemporal stability to be useful in capturing aspects of speech movement control that were not accessible with analytic techniques employed in earlier studies of speech motor control processes. The STI is proposed not as a replacement for traditional measures but as an additional analysis that provides, in a single value, information about the performer's composite output. The index is composite in that it reflects variability attributable to spatial and temporal aspects of control; it is also composite in the sense that variability over the entire movement trajectory is integrated into a single value.

This analysis is novel in some respects—for example, in that it may incorporate the waveform for an entire sentence—and it involves signal processing procedures that have not been standard in the speech production literature. Thus, the purpose of the present note is to summarize the rationale for speech movement trajectory analysis and to describe in more detail the methods used to compute the spatiotemporal index.

Rationale for Speech Movement Trajectory Analysis

The study of motor control processes has a rich heritage dating back to the pioneering work of Sherrington (1906). Thus, in the study of speech motor control, theoretical frameworks and analytic techniques can be borrowed and modified from investigations of many other kinds of movement systems. Other movement production systems, however, usually have an obvious unit of analysis—for example, the step or chewing cycle or a single aimed movement to a target. Speech, as the motor behavior conveying language, involves multiple linguistic units operating in parallel, and the appropriate unit for analysis has been a focus of debate since the earliest studies of speech production. Putative units have varied in size, from the feature to the syllable to the phrase (MacNeilage, 1970; Smith, 1992). From a motor control point of view it seems reasonable to define the analytic "window" in reference to movement units, and a single movement might be selected for analysis. However, it is well known that, because of effects of coarticulation, parameters of a single movement are affected by multiple components (e.g., segments, stress) of the utterance. There is no one-to-one correspondence between linguistic units and movements. Thus the size of the analytic window in speech kinematic analysis is not necessarily dictated by a dominant concept of "the unit" of behavior, and logically the size of the window, from single to multiple movements, depends on the specific experimental question being addressed.

Work from our laboratory has centered in recent years on the study of the development of normal speech in children and on the factors that lead to speech production disorders, such as stuttering. In the development of normal speech production skills and in the emergence of stuttering, it is hypothesized that different factors operating at many different levels affect speech motor processes. These factors range from linguistic variables, such as phonological encoding (e.g., Yaruss & Conture, 1996) or syntactic processing (i.e., Ratner, 1997), to emotional or autonomic variables (e.g., Smith & Kelly, 1997; Weber & Smith, 1990). Given all of these considerations, we were motivated to develop an analysis of speech kinematic output that could be used to examine the impact of variables operating over different processing levels and time scales, from single movements to movement sequences for an entire utterance.

Another issue, which led us to seek a more global or composite measure of overall performance, is that results of studies using measures at single points in time were often conflicting. Although statistical differences might be found for one measure, they might not be found for another. For example, in a simple, open-close lip movement sequence for a syllable such as *bob*, three standard measurements could be made (duration, peak velocity, and peak displacement) for both the opening and closing movements. Although such analyses are very useful in specifying spatial and temporal aspects of movement, we often get mixed results within and

between studies. Adams et al. (1993) provided a detailed review of the mixed results of studies using such measures to determine the effects of speaking rate changes on articulatory movements. For example, velocity of articulatory movements has been found to increase, decrease, or remain unchanged when speaking rate is increased. It is difficult to characterize performance as a whole on the basis of such results.

Therefore our goal was to develop a dependent variable that (a) is not tied to assumptions about specific underlying units, (b) provides a composite score that combines spatial and temporal aspects of performance, (c) can be used to probe the effects of more global variables (e.g., rate change or emotional arousal) as well as local variables (e.g., changing a single phoneme in an utterance), and (d) is amenable to application in data collected from young children.

Complementing this research on speech production, a review of work on limb motor control reveals a long tradition of spatial path (trajectory) analysis (Bullock & Grossberg, 1988; Georgopoulos, Kalaska, & Massey, 1981; Paulignan, MacKenzie, Marteniuk, & Jeannerod, 1991). Following the method of Georgopoulos et al. (1981), Paulignan et al. (1991) computed spatial paths of the wrist, thumb, and index finger motions during reaching and grasping cylindrical objects. Amplitude (non-normalized) variability of spatial paths was computed as a function of normalized time, and the cumulative sum of these standard deviations was used as a composite index of path variation. This type of analysis is attractive, because both the pattern of movement and its variability over time are revealed. Also this analysis could be applied to single speech movements or to a sequence of movements. Thus, this analysis met the criteria of not being tied to assumptions about specific units and was capable of being applied to movement trajectories with varying numbers of subcomponents. To date we have used this type of analysis (after some modifications, see below) on two-movement speech sequences (close-open lip movements perturbed by changing a single phoneme; Goffman & Smith, 1999) and on the multicomponent trajectories for a six-syllable utterance perturbed by rate change (Smith et al., 1995) or varying levels of utterance length and complexity (Maner, Smith, & Grayson, in press).

An additional step we took in the initial paper (Smith et al., 1995) was to incorporate a pattern-recognition procedure to determine whether a single, linearly scalable template was used across habitual, fast, and slow rate conditions. Clearly the results demonstrated that this was not the case. Although trajectories within a condition converged onto a single template, trajectories from the three rate conditions were sorted by the pattern-recognition algorithm with a high degree of accuracy. To our knowledge, this type of pattern-recognition analysis (Fukunaga, 1990) had not been used in earlier speech kinematic studies. Like the STI analysis the pattern-recognition procedures can be used on trajectories with variable numbers of movement components.

In summary, following a long tradition in limb motor control research, we developed a movement trajectory analysis that could be applied to single- or multiple-component speech movement trajectories. As we describe in detail below, we elected to not only time-normalize the trajectories but to amplitude-normalize them as well. With this process, we could determine the degree to which a basic, scalable movement template was being used across trials. The rationale for amplitude-normalization of movement trajectories is often explained by a handwriting example. Imagine a person is asked to write the word *dog* on a chalkboard and with pen on paper. The two handwriting trajectories would certainly have different durations and sizes, but time- and amplitude-normalization would reveal whether the two trajectories converged on a single, underlying pattern or template. Thus, implementation of this process would allow us to test how well, within a single condition, a set of trajectories converged onto a linearly scalable template. The STI represents the degree of fit. A perfect, linearly scalable pattern generator would produce a set of trajectories with an STI of zero (Smith et al., 1995). If the set of trajectories does not converge as well, the STI is high. In our work to date, the STI has proven to be reliable across studies and a useful composite index of motor performance.

Linear vs. Nonlinear Scaling

An issue that must be considered in designing a movement-trajectory analysis is whether linear or nonlinear scaling techniques should be used. The STI involves linear scaling of the displacement waveforms. In linear scaling all parts of the waveform are stretched or compressed by the same factor. In nonlinear scaling a warping function is defined that maps the data according to user-defined criteria. Nonlinear techniques are extremely powerful. A warping function could be created that would force any group of trajectories to converge onto a single template. Thus, the user has to define boundary conditions or rules for the warping function, and these are usually related to assumptions about what features of the signal are important.

Speech involves nonlinearities, and speech-recognition research on the acoustic signal routinely employs nonlinear algorithms—for example, to define the best exemplars of sound categories (Rabiner & Juang, 1993). Thus, early in our work we considered using nonlinear warping functions. We decided that, as a first step, we

would determine if classical, linear scaling procedures used in limb motor control worked to answer our experimental questions. We intended to use linear scaling only within conditions, because pilot work indicated that, if the target rate or prosody (e.g., stress) of the utterance is changed, the trajectories would contain nonlinear alterations across conditions. Also, we reasoned that the results would clearly indicate if linear methods were inadequate. For example, if after linear time-normalization, the waveforms did not tend to converge, the standard-deviation function would be extremely noisy. The cumulative sum of these, the STI, would be meaningless, and group and condition effects would be inconsistent. In fact, the movement trajectories do tend to converge, and reliable group and condition effects have been found. Such results imply that, when the target utterance is unchanged, repeated productions of the utterance are characterized by linearly scalable waveforms. The degree to which they do not scale linearly, represented by the STI, tells us something meaningful about the composite performance of the speaker. The experimental questions driving research in our laboratory (e.g., how syntactic complexity and/or the length of the utterance affect speech motor stability) seem to be well served by this simple, linear analysis.

Lucero et al. (1997) describe a nonlinear method for scaling speech movement waveforms. They employed the same utterance used by Smith et al. (1995) and compared the linear time-normalization technique to the nonlinear method. They concluded that the nonlinear warping function is better for preserving important landmarks on the waveforms, so that the resulting average of a set of trajectories (they used acceleration rather than displacement trajectories) is a truer representation of the underlying pattern of behavior. Visual inspection of the acceleration records they include suggests that this is true, and their technique of nonlinear time normalization seems very promising. However, Lucero et al.'s only stated goal is to find a technique that produces the best average of a set of acceleration records—one that preserves timing and amplitude of peaks. It is not clear why the acceleration peaks are assumed to be the most significant features of speech movement dynamics, as the authors do not address the theoretical or experimental advantage of preserving peaks in the acceleration-time functions. Ultimately nonlinear scaling, because rules for the scaling algorithm must be created, requires the experimenter to decide which features of the signal are most significant or are hypothesized to be most closely constrained by the operation of underlying units. This is an important experimental question, one that returns us to the issue of what units are operating in speech production.

The choice of linear versus nonlinear scaling techniques ultimately must be decided on the basis of the experimental question. We explicitly did not wish to design our analysis to optimize the preservation of landmarks assumed to be most fundamentally related to underlying units, because we were aligned theoretically to the position that particular units do not have "preferred" status. Thus we opted to try simple, linear techniques so that we did not have to decide which were the most significant aspects of the displacement trajectory. Thus in Smith et al. (1995), following linear time-normalization, we averaged the set of waveforms for an utterance spoken at each of three different rates. These average templates were then assumed to represent the pattern of behavior for that condition and were used to sort the input waveforms into rate categories. The worst performance of the pattern-recognition procedures was 96% accuracy in sorting. Thus it seems that this linear method of representing the pattern of the set of displacement trajectories was very successful. In summary, both linear and nonlinear methods of normalizing sets of trajectories hold promise for speech-production research. Each experimenter must decide which methods best suit his or her experimental questions.

Methodological Issues in Movement-Trajectory Analysis

Extraction of Records for Analysis

A general goal of the STI and other movement-trajectory analyses is to search for convergence of a set of trajectories onto a common template. Thus a set of trajectories for each experimental condition is required. In the STI procedure the standard deviation is computed as a function of normalized time for the multiple trials in each condition. Initially, we used 15 but later found that 10 trials per condition produced similar results and reduced the time required for data collection. Minimizing data collection time is important, because much of our work involves testing young children.

After collection, the kinematic signal streams are imported into an analysis program (Matlab, Mathworks, 1994). Before further analysis the displacement records are digitally low-pass filtered to remove noise. Then, using standard kinematic analysis procedures, velocity is computed using the three-point difference method (Wood, 1982). As the next step, the trajectories must be selected to enter into the STI and/or pattern-recognition analysis. Start and end points must be reliably selected for the movement trajectory for each trial. The number of movement components in the trajectories depends upon the question under investigation. For example, the effects of changing a single phoneme in an utterance were examined by extracting close-open movement sequences for lip movement into and out of a consonant (Goffman & Smith, 1999). Following standard

procedures, movement onsets and offsets were identified as the point at which the velocity signal changed sign (e.g., onset of closing movement is the point at which velocity becomes positive). In other experiments, effects of rate change and subject's age on the stability of movements for a phrase were determined by extracting lip movement for the entire utterance, from the points of peak velocity of the first and last movements for the phrase (Smith et al., 1995; Smith & Goffman, 1998).

To implement these procedures, displacement and velocity signals are simultaneously displayed within a Matlab program. The experimenter then selects the approximate location of the velocity event by visual inspection (e.g., a peak or a zero-crossing). Within a predetermined window of points (e.g., 21 points), an algorithm is then employed that precisely selects the desired point. The latter algorithm is important, as it significantly reduces error in end-point selection.

In summary, standard kinematic landmarks are reliably selected to extract the record that is to be used for the STI computation. One advantage of this analysis is that there is only one assumption involved: that records containing the same number of movements are reliably extracted for each experimental condition. Beyond this step, there are no assumptions about which points in the signals are more significant than others, and the entire trajectory for each trial is entered into the analysis.

Amplitude- and Time-Normalization

At this point, there is a set of movement trajectories (generally 10 or 15) for each condition (this analysis could be applied to displacement-time or any higher derivative-time function). To determine the degree to which the set of waveforms converges onto a single template, the set of waveforms is linearly amplitude- and time-normalized. Linear amplitude normalization is trivial; it is simply the computation of the *z* score (calculate the mean and standard deviation of each displacement record, then subtract the mean and divide by the *SD*).

The resulting amplitude-normalized trajectories for a given condition vary in the number of points they contain because of variation in the overall duration of each trial. For example, a set of displacement signals extracted according to the procedures outlined above for the phrase "buy bobby a puppy" produced at habitual rate by a normal adult speaker could range from 240 to 270 points (with a sample rate of 250 samples/s). The next step in the trajectory analysis is linear time-normalization. As in earlier studies, the motivation for this step is to determine if, after variation attributable to overall duration is removed, the set of trajectories converge onto a single pattern (e.g., see Figure 5 of Gutfreund et al.,1996). In this step each waveform is mapped onto a constant number of points. In the limb motion literature, this procedure traditionally has been computationally trivial. Each trajectory for multiple trials of a single movement is simply time-normalized by taking 100 time "slices" (Georgopoulos et al., 1981; Paulignan et al.,1991). This results in a set of trajectories that are each 100 points in length (usually depicted as 0–100% in relative time). This procedure works well for simple, single-movement data records of the limb studies. This procedure would not work well for records with multiple movements, as information about movement components and shape potentially would be lost. Also note that records cannot simply be cut to the length of the shortest record in the set, as this process would result in the loss of movement components for the longer records.

Therefore, a more sophisticated method was required to map variable-length records onto a constant number of points. This process—interpolation—is a standard procedure in signal processing, and it may be accomplished through a variety of means (Conte & de Boor, 1980; Proakis & Manolakis, 1995). We elected to use trigonometric interpolation. In this method a discrete Fourier transform is computed to obtain the weighting coefficients for a polynomial function that is then used to map the data onto a constant number of points (we have used 1,000 points, but for trajectories with fewer components, fewer points would be adequate). In the Appendix, the details of the trigonometric procedure and a comparison of its performance with that of other standard techniques for interpolation, including the cubic spline (e.g., Lucero et al. 1997), are included. As illustrated in the Appendix, the trigonometric interpolation procedure of Smith et al. (1995) results in no significant loss of signal information. For readers not interested in the level of detail reported in the Appendix, the following summary presents the salient points.

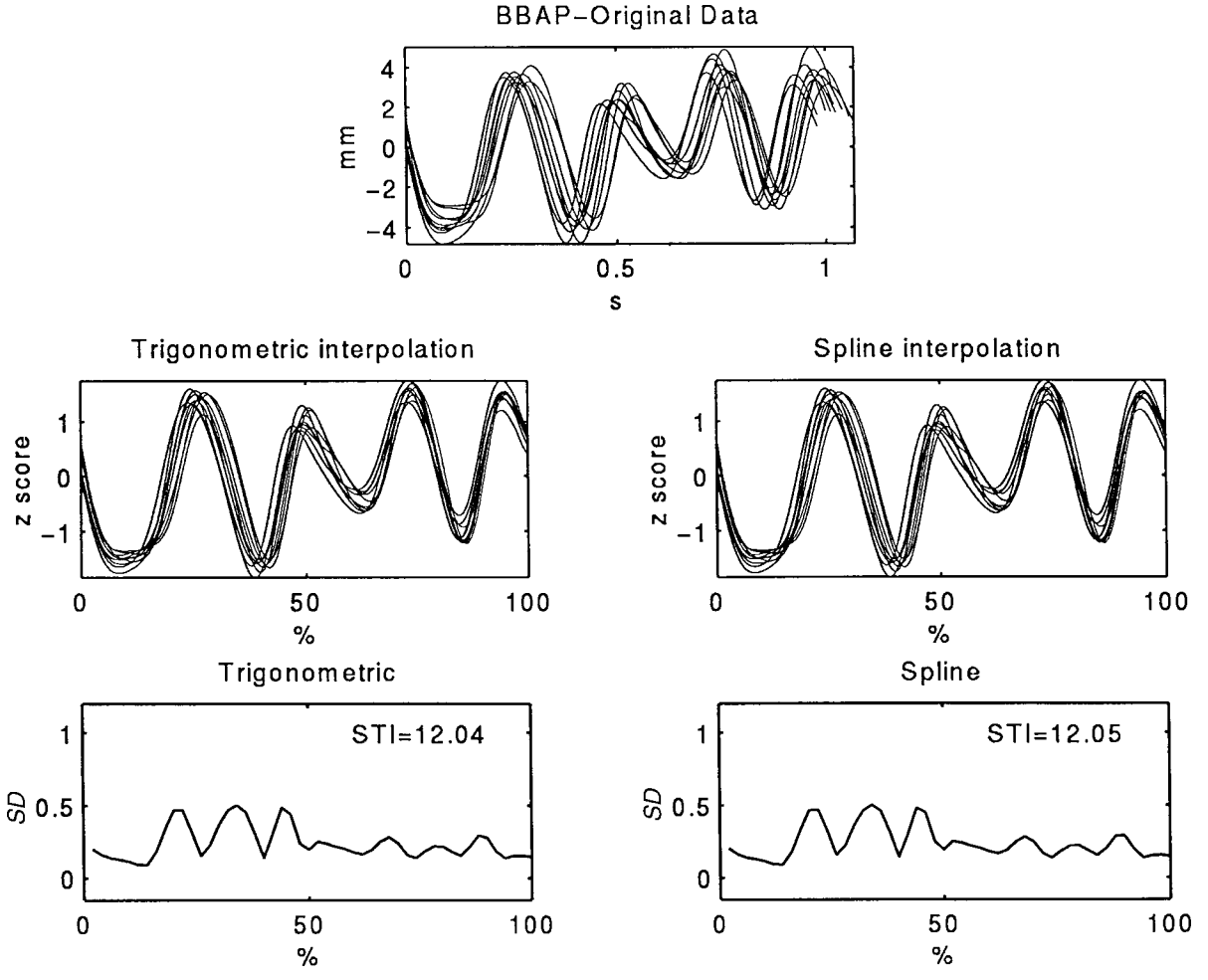
Any interpolation process produces some level of error; however, on modeled data, the mean square error with the interpolation techniques we compared (trigonometric, spline, and resampling) was extremely small—less than 0.4% for all three (see Table B, Appendix). We preferred the trigonometric method, because mathematically it is simpler and in many ways more elegant. Instead of using piece-wise polynomials, the basis functions for trigonometric interpolation are continuous (Conte & de Boor, 1980; Lanczos, 1956). The cubic spline (de Boor, 1978), in which a separate polynomial function is computed between each pair of adjacent points in the record, is computationally slower. We would recommend that the experimenter beginning to implement this type of analysis for speech kinematic

data use the spline interpolation, because, as outlined below, for many users, the spline may be simplest to implement. Both methods produce negligible error and are standard Matlab functions (interpft [an end-point correction is needed, see Appendix] and interp1, which implements a spline method).

The trigonometric interpolation process uses a Fourier transform. The Fourier analysis is completed, not to filter the data, but to obtain weighting coefficients for the interpolation function. Concern has been expressed (Lucero et al., 1997, Footnote 1) that, for the slow condition of normal adult speakers in Smith et al. (1995), the trigonometric interpolation with coefficients derived from 10 harmonics excessively filters the data and results in the loss of signal information. This is not the case. As the analysis in the Appendix indicates, even for the slowest speaker in that study (slow condition records ranged from 2.2 to 2.4 s), 98–99% of the signal energy was contained in the first 10 harmonics. Longer displacement records do require more coefficients to represent 99% of the signal energy. In a later study (Smith et al., in progress), with adults who stutter speaking at a slow rate, records greater than 4 s in length were obtained. In such cases we have found that coefficients from 40 harmonics are needed in the trigonometric interpolation to represent all of the energy in the signal (see Table A-1, Appendix). As a rule, the practice we have used seems reasonable: Employ enough coefficients in the interpolation function to represent 99% of the energy in the signal.

The conclusion we draw from this discussion and the analysis included as an Appendix is that the trigonometric and spline interpolations are functionally equivalent for linear time-normalization of kinematic waveforms such as those we have studied. For a straightforward confirmation of this point, Figure 1

Figure 1. Comparison of trigonometric and spline methods for time normalization in the STI analysis. Amplitude normalization is the same for both. Top plot: original displacement waveforms for 10 repetitions of "buy Bobby a puppy." Left middle plot: time normalization with the trigonometric procedure. Right middle plot: time normalization with the spline procedure. Bottom plots: standard deviation as a function of relative time for the data normalized with each procedure.



compares the results of the STI analysis using spline and trigonometric interpolation. The trigonometric method requires the end-point correction procedure (see Appendix) and an understanding of how many coefficients are necessary to capture all the information in signals of varying length. Therefore, the burden on the user is greater. The spline interpolation is accomplished with a single line of Matlab code, and although it is computationally more burdensome, the burden is borne by the computer rather than the user.

Calculation and Interpretation of the STI

At this point in the analysis, there is a set of waveforms (usually 10 or 15), each composed of 1,000 points, for each condition. As a simple first step in assessing how well the waveforms converge after normalization, the set of normalized records can be compared visually to the original data (e.g., compare the upper and middle panels of Figure 1). Some degree of convergence is almost always apparent. To quantify this degree of convergence, the standard deviation of the 10 or 15 points at 2% intervals in relative time is calculated (every 20th point for a 1,000-point normalized time base). The decision to use 2% rather than 1% intervals was arbitrary. We decided that 50 *SD* values were adequate, but 100 could be used. The STI, which is the sum of these 50 *SD*s, is also therefore an arbitrary number. If we used 1% intervals, STI values would approximately double.

The precise value of the STI is also utterance specific, and direct comparisons across utterances cannot be made. For example, within a subject, STI values for trajectories containing only two movements are lower than those containing more movement components. The standard deviation tends to be higher when velocity of movement is high; thus if there are more intervals of high velocity movement in the record, the STI will tend to be higher. The STI also depends on the degree of constraint of movements of the specific articulator under study. In pilot work we have found that STIs are intrinsically higher for lip motion when the utterance does not contain segments with labial targets. For example, within a subject the STI computed for lip motion for "I like that tiny cat" will be higher than that for "buy Bobby a puppy" or "mommy bakes pot pies."

The STI across studies has been remarkably reliable for a single utterance ("buy Bobby a puppy") spoken at habitual rate and loudness by normal young adults. Over a period of 5 years, we have tested five small groups (8–10 subjects) of young adults and have found ranges of 12.4–14.2 for the mean STI and 2.3–2.7 for the *SD*. In a study now in progress, we have tested 30 young adults on this utterance and have found a mean STI of 12.7 and $SD = 3.1$.[1] We do not know of another kinematic speech measure that is equally reliable.

The value of the STI analysis, therefore, lies in providing a simple, composite score that can be used to compare stability and patterning across subject groups and within subjects across conditions. Returning to the goals that initially motivated this analysis, it seems that the STI approach has been successful in many ways, but there are, of course, limitations. The STI has been used to study trajectories of single fleshpoints over time. In work in progress, we have applied the analysis to interarticulator variables, such as lip aperture (upper lip–lower lip signal) and relative phase angle data (upper lip phase relative to lower lip phase). Preliminary data suggest that the normalization method is also useful for capturing aspects of interarticulator coordination over time. In conclusion, recently introduced methods (including linear and nonlinear approaches) for normalizing trajectories of single-effector movements or interarticulator phase trajectories seem to hold great promise for research in which speech movement output is analyzed for clues concerning the underlying neural control processes.

Acknowledgments

This work was supported by National Institutes of Health (NIDCD) grants DC00559 and DC02527. Our colleague Claire McGillem passed away during the final revision of this paper. We acknowledge with deep gratitude Claire's long-term and insightful contributions to our work.

[1]In relation to the issue of reliability across groups of normal speakers, we are sometimes asked if we carefully instruct participants on how to speak and if we exclude large numbers of volunteers. Just as with most speech measures, if subjects do not produce the target phrase (e.g., they change loudness, rate, or other aspects of the utterance across trials when not instructed to do so), the STI will be affected. Potential participants, however, are rarely excluded for this reason. We do not give any instructions on how to speak other than a scaling procedure (e.g.,"speak twice as fast as your normal rate") for rate or loudness changes. As an example, in a recent study of young adults, a total of 34 volunteers were recruited, and 4 were excluded from the study. Three did not pass the language-screening tool. One was excluded because she could not reliably produce the longer sentences in the protocol. It was not necessary to exclude anyone because he or she produced uninstructed changes in prosodic characteristics of the utterances. Another potentially important experimental variable affecting reliability is the order of sentence production. In most studies we have asked participants to produce the sentences in blocks of approximately 5 to 10 repetitions, and we collect the data in 20- to 30-s trials. In one study with seven sentence stimuli, utterances were produced in a semirandom order, so that kinematic data from only one utterance were collected per trial. We found that the STI for the standard utterance we have used ("buy bobby a puppy" at normal rate and loudness) from that study ($X = 13.6$, $SD = 2.4$) was not different from the values produced when the repetitions are blocked together. Thus the random stimulus presentation, although preferable in terms of active language processing demands of the task, did not apparently make a difference in the stability of motor execution for normal adult speakers.

References

Ackermann, H., Hertrich, I., & Scharf, G. (1995). Kinematic analysis of lower lip movements in ataxic dysarthria. *Journal of Speech and Hearing Research, 38,* 1252–1259.

Adams, S. G., Weismer, G., & Kent, R. D. (1993). Speaking rate and speech movement velocity profiles. *Journal of Speech and Hearing Research, 36,* 41–54.

Atkeson, C. G., & Hollerbach, J. M. (1985). Kinematic features of unrestrained vertical movements. *Journal of Neuroscience, 5,* 2318–2330.

Bullock, D., & Grossberg, S. (1988). Neural dynamics of planned arm movements: Emergent invariants and speed-accuracy properties during trajectory formation. *Psychological Review, 95*(1), 49–90.

Conte, S. D., & de Boor, C. (1980). *Elementary numerical analysis.* New York: McGraw Hill.

de Boor, C. (1978). *A practical guide to splines.* New York: Springer-Verlag.

Flash, T., & Hogan, N. (1985). The coordination of arm movements: An experimentally confirmed mathematical model. *The Journal of Neuroscience, 5,* 1688–1703.

Fukunaga, K. (1990). *Introduction to statistical pattern recognition* (2nd ed.). Boston, San Diego, New York: Academic Press Inc. Harcourt Brace Jovanovich, Publishers.

Georgopoulos, A. P., Kalaska, J. F., & Massey, J. T. (1981). Spatial trajectories and reaction times of aimed movements: Effects of practice, uncertainty, and change in target location. *Journal of Neurophysiology, 46,* 725–743.

Goffman, L., & Smith, A. (1999). Development and phonetic differentiation of speech movement patterns. *Journal of Experimental Psychology: Human Perception and Performance, 25,* 649–660.

Gutfreund, Y., Flash, T., Yarom, Y., Fiorito, G., Segev, I., & Hochner, B. (1996). Organization of octopus arm movements: A model system for studying the control of flexible arms. *Journal of Neuroscience, 16,* 7297–7307.

Kent, R. D., & Moll, K. L. (1975). Articulatory timing in selected consonant sequences. *Brain and Language, 2,* 304–323.

Kleinow, J., & Smith, A. (in press). Influences of length and syntactic complexity on the speech motor stability of adults who stutter. *Journal of Speech, Language, and Hearing Research.*

Kuehn, D. P., & Moll, K. L. (1976). A cineradiographic study of VC and CV articulatory velocities. *Journal of Phonetics, 4,* 303–320.

Lanczos, C. (1956). *Applied analysis.* Englewood Cliffs, NJ: Prentice Hall.

Lucero, J. C., Munhall, K. G., Gracco, V. L., & Ramsay, J. O. (1997). On the registration of time and patterning of speech movements. *Journal of Speech, Language, and Hearing Research, 40,* 1111–1117.

MacNeilage, P. F. (1970). Motor control of serial ordering of speech. *Psychological Review, 77,* 182–196.

Maner, K., Smith, A., & Grayson, L. (in press). Influences of utterance length and complexity on speech motor performance in children and adults. *Journal of Speech, Language, and Hearing Research.*

The Mathworks, Inc. (1994). *Matlab: High-Performance Numeric Computation and Visualization Software* (Version 4.0) [Computer Software]. Natick, MA: Author.

Ostry, D. J., Cooke, J. D., & Munhall, K. G. (1987). Velocity curves of human arm and speech movements. *Experimental Brain Research, 68,* 37–46.

Paulignan, Y., MacKenzie, C., Marteniuk, R., & Jeannerod, M. (1991). Selective perturbation of visual input during prehension movements. *Experimental Brain Research, 83,* 502–512.

Proakis, J. G., & Manolakis, D. G. (1995). *Digital signal processing.* Englewood Cliffs, NJ: Prentice Hall.

Rabiner, L., & Gold, B. (1975). *Theory and application of digital signal processing.* Englewood Cliffs, NJ: Prentice Hall.

Rabiner, L., & Juang, B. (1993). *Fundamentals of speech recognition.* Englewood Cliffs, NJ: Prentice Hall.

Ratner, N. B. (1997). Stuttering: A psycholinguistic perspective. In R. F. Curlee & G. M. Siegel (Eds.), *Nature and treatment of stuttering: New directions* (pp. 99–127). Boston: Allyn & Bacon.

Shaiman, S., Adams, S. G., & Kimelman, M. D. (1997). Velocity profiles of lip protrusion across changes in speaking rate. *Journal of Speech, Language, and Hearing Research, 40,* 144–158.

Sherrington, C. S. (1961). *The integrative activity of the nervous system* (2nd ed.). New Haven: Yale University Press. (Original work published 1906)

Smith, A. (1992). The control of orofacial movements in speech. *Critical Reviews in Oral Biology and Medicine, 3*(3), 233–267.

Smith, A., & Goffman, L. (1998). Stability and patterning of speech movement sequences in children and adults. *Journal of Speech, Language, and Hearing Research, 41,* 18–30.

Smith, A., Goffman, L., Zelaznik, H. N., Ying, G., & McGillem, C. M. (1995). Spatiotemporal stability and patterning of speech movement sequences. *Experimental Brain Research, 104,* 493–501.

Smith, A., & Kelly, E. (1997). Stuttering: A dynamic, multifactorial model. In R. Curlee & G. Siegel (Eds.), *Nature and treatment of stuttering: New directions* (pp. 204–217) Boston: Allyn & Bacon .

Weber, C. M., & Smith, A. (1990). Autonomic correlates of stuttering and speech assessed in a range of experimental tasks. *Journal of Speech and Hearing Research, 33,* 690–706.

Wohlert, A. B., & Smith, A. (1998). Spatiotemporal stability of lip movements in older adult speakers. *Journal of Speech, Language, and Hearing Research, 41,* 41–50.

Wood, G. A. (1982). Data smoothing and differentiation processes in biomechanics. *Exercise and Sports Science Review, 10,* 308–362.

Yaruss, S. J., & Conture, E. G. (1996). Stuttering and phonological disorders in children: Examination of the covert repair hypothesis. *Journal of Speech and Hearing Research, 39,* 349–364.

Zimmermann, G. N. (1980a). Articulatory dynamics of fluent utterances of stutterers and nonstutterers. *Journal of Speech and Hearing Research, 23,* 95–107.

Zimmermann, G. N. (1980b). Articulatory behaviors associated with stuttering: A cinefluorographic analysis. *Journal of Speech and Hearing Research, 23,* 108–121.

Received July 28, 1998

Accepted July 6, 1999

Contact author: Anne Smith, PhD, Audiology and Speech Sciences, Heavilon Hall, Purdue University, West Lafayette, IN 47907-1353. Email: asmith@purdue.edu

Appendix. Linear Time-Normalization.

The goal of linear time normalization is to transform an n-point representation of a continuous-time function f(t) into an m-point representation of the same function, with minimal error. If the underlying function f(t) is known, this process is trivial and can be accomplished with a direct scaling of the time axis:

$$f_1(t) = f(nt/m),$$

where

n = the number of points in the original sampled representation

m = the number of points in the transformed representation

f = the original known function

f_1 = the linearly time-normalized function

If f(t) is unknown, however, it must be estimated from the n-point sampled representation and then the above formula used to generate an equivalent m-point representation. This process is known as interpolation and may be accomplished by a variety of methods, including polynomial interpolation, piecewise polynomial interpolation (e.g., splines), trigonometric interpolation, and resampling techniques, among many others.

The primary goal of interpolation is to minimize error, defined at each point x as the difference between the original function at f(x) and the interpolating function, which we will denote p(x). Again, without knowledge of the original function, the error cannot be calculated directly. However, if we can make some assumptions about this function (for example by assuming a bound on the second or third derivative in piecewise polynomial approximation), we can bound the error using a technique such as a Taylor Expansion of the interpolating function. Error is affected by a number of factors—primarily the complexity of the function used for interpolation, the length of the time axis under any single interpolating function, and the spacing of the time axis points. (Spacing the interpolating points as Chebyshev points rather than linearly spaced points, for example, can be shown to minimize the error of an n-degree polynomial interpolant.) We will examine several predominant interpolation methods and compare their advantages and disadvantages. It should be noted than any of the methods used here, when implemented properly, will result in extremely small error—typically less than 1% distortion.

Piecewise Polynomial Interpolation

Interpolation using piecewise polynomials is perhaps the most popular interpolation method. It consists of estimating a separate polynomial interpolating function between adjacent points on the time axis. Each polynomial is determined by a least-squared error minimization combined with constraints on end-point continuity of the polynomials and frequently first and second derivative continuity as well. Because all polynomial representations are mathematically equivalent (de Boor, 1978), the polynomial basis used may be chosen as desired. Splines, more specifically cubic B-splines, are often the representation of choice because of certain mathematical advantages of the coefficient structure. The error of this approach is determined by the square of the point spacing and the extrema of the third derivative of the underlying function, if known. Splines are a very precise method, but they have a high degree of complexity and they are computationally time consuming.

Resampling

Resampling is a technique used primarily by dedicated signal processing systems, which are designed to handle the filtering needs of this method. To resample a sampled signal (in our case the n-point representation of the function) by a rational factor I/D, where I and D are integers, requires the following steps:

a. Upsample the signal by the factor I.

b. Filter—to remove the high frequency aliasing from step (a).

c. Downsample the signal by the factor D.

The error of this technique, when considered from an interpolation perspective, is related to the filter in step (b). It is also assumed that the original signal has been prefiltered by an appropriate band-pass filter. This assumption ensures perfect reconstruction if the filter in (b) is ideal. Usually a polyphase filter is employed, where the number of subfilters is at least the upsampling factor I. Because of the filter complexity, this method would normally not be recommended (see Proakis & Manolakis, 1995) for high integers I and D. In the case of the kinematic data in our research, for example, to linearly time-normalize a 239-point sample to 1,000 points would produce I = 1,000, D = 239, requiring a minimum of 1,000 subfilters for good performance. The method is primarily designed for continuous resampling as opposed to resampling of small sections of data, such as those extracted in kinematic analysis.

Trigonometric Interpolation

Trigonometric interpolation is another common technique for performing time normalization (Lanczos, 1956). It has been used for many years in the signal processing community because of its simplicity when compared with such techniques as piecewise polynomial approximation. Another advantage of

trigonometric interpolation is that the interpolating function is a well-behaved analytic expression upon which further mathematical analysis, such as integration and differentiation, can be performed. In this method, the original function f(t) is represented using trigonometric basis functions rather than polynomials. The N-point representation of the original signal is decomposed into complex sinusoidal coefficients using the Discrete Fourier Transform (DFT).

These coefficients may then be used to reconstruct the signal along any time axis desired. In both trigonometric and resampling methods, any error will be concentrated around points of discontinuity and derivative discontinuity, because of Gibbs's phenomena (Lanczos, 1956). For a finite section of a band-pass filtered waveform, the only points of this nature are the end-points (see the following section for minimizing end-point error). Further, in trigonometric interpolation, if the number of coefficients is reduced, the speed of interpolation is increased proportionately, at the expense of some slight increase in error. Note that when using a DFT in this manner, the goal is interpolation, not analysis of the frequency content of the signal, which is much better accomplished using a more accurate estimator of a signal's power spectral density, such as the Welch or Blackman-Tukey methods (Rabiner & Gold, 1975).

The trigonometric technique explained here is the method we chose for performing linear time-normalization (Smith et al., 1995), in a large part because of its simplicity. We used only the first few coefficients, because more than 99.9 % of signal and window energy was contained therein. We also added an additional technique to minimize the error as described below.

It can be shown that an N-point DFT is equivalent to calculating the discrete-time Fourier series of an infinite periodic signal with period N (formed by replicating the original N-point signal). The number of coefficients required for adequate representation of the signal is related not only to the energy contained at various frequencies of the original signal, but also to the size and placement of the N-point window and any possible discontinuities (and to a lesser extent, derivative discontinuities) between the signal at its left and right end-points. To improve performance, the endpoints may be forced to match by removing a linear trend (i.e., by subtracting a line going through the first and last points) from the original signal before interpolation and re-adding it after interpolation. This is effectively the addition of two additional coefficients representing linear rather than trigonometric components.

To demonstrate the impact of the number of coefficients employed in trigonometric interpolation, in Table A-1 we have evaluated examples of actual kinematic data using increasing numbers of coefficients. The additional error attributable to limiting the coefficients (excluding interpolation error, which is discussed in the next section) is seen as lost signal energy and may be quantified by showing the percentage of total signal energy contained in the coefficients used. The number of total coefficients required to obtain 99% of the signal energy is greater for longer data record lengths. Results (Table A-1) show that relatively few coefficients are needed to reconstruct the signal without significant loss of energy.

Table A-1. The number of coefficients needed to represent 99% of the energy in the signal increases with the length of the record.

Number of coefficients	Duration of input waveforms (with sample rate = 256)			
	0.26 s	**1.04 s**	**2.33 s**	**4.81 s**
3	.999	.111	.209	.172
5	—	.933	.907	.639
10	—	.999	.987	.972
20	—	—	.999	.997
40	—	—	—	.999

Example

To see how similarly the various techniques for linear interpolation perform, we have constructed a simple example, in this case using a known function, so error can be accurately determined. The function is plotted in Figure A-1; it is (1+0.5*sin(2*pi.*x./280)).*sin(2*pi.*x./80).

This 271-point function was interpolated to 1,000 points in length with each method. The interpolations were done using the Matlab functions interpft, interp1, and resample. Interpft was modified to use a fixed number of coefficients and to include the detrending step described above to control end-point distortion. As shown in Table A-2, the spline approach has the least error of the group, but the maximum mean square error for all of these methods (.4%) is negligible and unlikely to have significant impact on calculations such as the STI (see Figure 1 in the main text).

Figure A-1. Plot of the known function generated to estimate interpolation error.

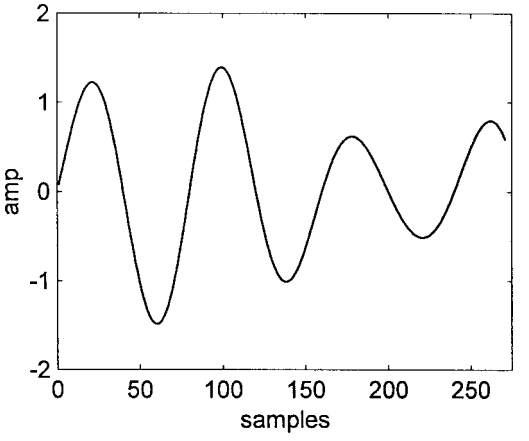


Table A-2. Mean Square Error (MSE) after interpolation of a known function using the various methods (271 point record interpolated to 1,000 points).

Method	MSE
Cubic Spline	1×10^{-13}
Resampling	3.7×10^{-3}
Trigonometric (10 coeff.)	4.0×10^{-3}
Trigonometric (40 coeff.)	3.2×10^{-3}