Comparison of Measures of Variability of Speech Movement Trajectories Using Synthetic Records

RESEARCH NOTE

Jorge C. Lucero
University of Brasilia,
Brazil

In speech research, it is often desirable to assess quantitatively the variability of a set of speech movement trajectories. This problem is studied here using synthetic trajectories, which consist of a common pattern and terms representing amplitude and phase variability. The results show that a technique for temporal alignment of the records based on functional data analysis allows us to extract the pattern and variability terms as separate functions, with good approximation. Indices of amplitude and phase variability are defined, which provide a more accurate assessment of variability than previous approaches.

KEY WORDS: variability, speech movements, time normalization, functional data analysis

This work deals with the problem of extracting the common pattern and variability from a set of speech movement records. It is a follow-up from a previous article (Lucero, Munhall, Gracco, & Ramsay, 1997) in which a new technique based on functional data analysis (FDA; Ramsay & Silverman, 1997) was applied to extract the pattern from such a set. That technique was based on performing an optimal nonlinear transformation of the time scale (nonlinear time normalization) so as to align the trajectories in time. The common pattern was computed as the average of the normalized trajectories, and the amplitude variability of the set was visualized by computing the difference of each normalized trajectory from the average. Further, the computed time transformations were considered as a representation of the phase variability of the set.

In Lucero et al. (1997), we concluded that the technique offered some advantages over previous approaches for studying patterns and variability (e.g., Smith, Goffman, Zelaznik, Ying, & McGillem, 1995). We claimed that the computed average was visually closer to the actual pattern of the trajectories than the average computed using other techniques (such as the linear normalization used by Smith et al., 1995). Also, the technique allowed us to extract both amplitude and phase variability of the set as separate components, and to visualize the distribution of such variability along the trajectory length. However, these conclusions were based only on visual inspection of the results. The question of how well the extracted patterns and variability approximated the actual pattern and variability of the set was left unanswered.

Similar problems have been considered in measures of signal-to-noise ratio of speech signals (e.g., Neilson & O'Dwyer, 1984; Strick & Boves, 1991). One such measure for voice signal irregularity is the harmonics-to-noise ratio (HNR; Yumoto, Gould, & Baer, 1982), defined as the ratio of the harmonic energy to the noise energy of the voice signals. The harmonic component of the signals is computed as the common pattern that repeats cycle to cycle and the noise component as the difference of each cycle to the harmonic component. It has been shown (Qi, 1992; Qi, Weinberg, Bi, & Hess, 1995) that an accurate computation of the HNR requires a nonlinear normalization of the individual cycles. Following those results, the above FDA technique was applied to compute the HNR of synthetic and recorded signals (Lucero & Koenig, 2000), with the advantages noted above. More recently, FDA nonlinear normalization has been successfully used to assess variability of glottal gestures in children versus adults (Koenig & Lucero, 2002; Koenig, Lucero, & Löfqvist, 2003) and to compare variability of simultaneous articulatory movements under different phonetic conditions (Koenig et al., 2003; Lucero & Löfqvist, 2003).

It might be useful to consider in more depth the extraction of patterns and variability using FDA because it is an interesting tool for studying aspects of voice and speech production. Such study must be based on a mathematical model of the signals and a clear definition of the terms involving patterns and variability. Following this idea, the current work presents a numerical experiment to assess the accuracy of the FDA normalization technique. A simple mathematical model will be used to generate trajectory waveforms, with known components of variability. FDA nonlinear normalization will then be used to assess how well it predicts the actual variability of the trajectory set.

Previous approaches based on linear (Smith et al., 1995; Smith, Johnson, McGillem, & Goffman, 2000) and piecewise linear (Ward & Arnfield, 2001) normalization will also be tested for comparisons with the FDA technique. Linear normalization consists of linearly stretching the time scale of the trajectories to a common length. Although this technique eliminates length (duration) variation between the trajectories, variation in the timing (phase) of events within the trajectories still persists (e.g., see Figure 3 in Lucero et al., 1997). Such phase variations could be due to variations in speaking rate, insertion of pauses, phrase boundaries, or similar effects. Thus, variability measures based on this approach are composite indices of both amplitude and phase variability. Piecewise linear normalization attenuates the effect of phase variability by breaking the trajectories into intervals defined by specific key events, which are prealigned first. Next, linear normalization is applied within the intervals. Although any phase variations at the key events are removed by the prealignment, such variations still remain within each interval. Those variations will still influence the measure of variability, although in smaller degree than measures using linear normalization. It might be helpful to have separate measures of amplitude and phase variability for a better assessment of speech movement variations. In this way, variations produced by speaking rate, pause insertion, and other temporal effects could be analyzed separately from variations in the shape or amplitude of the movements.

Synthesis of Speech Movements

Speech movements with characteristics similar to those used by Lucero et al. (1997) and Smith et al. (1995) are considered: the vertical displacement of the lower lip (in its combined motion with the jaw) during the repeated production of the sentence "Buy Bobby a puppy." A simple way to produce repeated trajectories with a common pattern is to assume that they are generated by a clock running at a central frequency ω_p. Variations of this modeling technique have been used in studies on voice perturbation measures (Lucero & Koenig, 2000; Titze & Liang, 1993) to generate waveforms of known added variability.

In the absence of any variability, the clock produces a pattern $p(\theta)$ at each cycle, where $\theta \in [0, 1]$ is a measure of the clock's phase, relative to the cycle length (we consider this measure instead of the usual angular phase $\in [0, 2\pi]$, to avoid carrying a factor of 2π and thus simplify the notation). The following expression is adopted for the pattern:

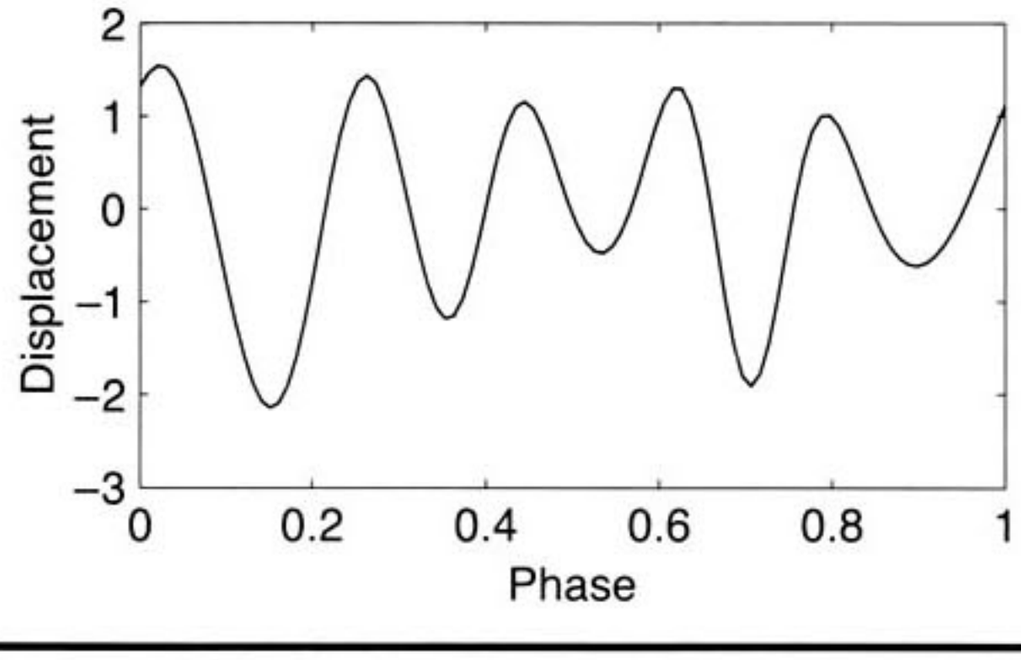
$$p(\theta) = \sum_{n=1}^{10} \Re(C_n)\cos(2\pi n\theta) + \Im(C_n)\sin(2\pi n\theta) \qquad (1)$$

where $\Re(C_n)$ and $\Im(C_n)$ are the real and imaginary parts, respectively, of coefficients in Table 1. These values were obtained by extracting the 2nd to 11th Fourier coefficients of a lip movement displacement during the

Table 1. Fourier coefficients for Equation 1.

n	C_n
1	$-0.053 - i0.068$
2	$0.349 - i0.224$
3	$0.260 - i0.193$
4	$0.564 + i0.413$
5	$0.594 + i0.693$
6	$-0.404 - i0.412$
7	$-0.070 + i0.120$
8	$0.085 + i0.095$
9	$0.016 - i0.063$
10	$-0.023 + i0.026$

Figure 1. Pattern $p(\theta)$ used to generate speech trajectories.



production of the above sentence at normal speed (Lucero et al., 1997). The first coefficient is not used, to produce a pattern with zero mean. Also, the magnitude of the remaining coefficients was scaled to produce a trajectory with $SD = 1$. The waveform of $p(\theta)$ is shown in Figure 1.

The pattern is produced with the clock running at a constant speed ω_p. Each cycle i of the clock produces a trajectory $x_i(t) = p[\theta(t)]$, with $d\theta/dt = \omega_p$, $t \in [0, T_p]$, $\theta(0) = 0$, $\theta(T_p) = 1$. To simplify the explanation, and without loss of generality, we consider dimensionless quantities and set $\omega_p = 1$ and $T_p = 1$.

We assume now that the speed of the clock has some variability, introduced by variations in the timing of individual movements of the lips. Such variations would be consequence of, for example, variations in speaking rate, insertion of pauses, or phrase boundaries. At each cycle i, the speed is

$$\frac{d\theta_i}{dt} = 1 + \gamma_i(t) \tag{2}$$

where $\gamma_i(t)$ is a random function with zero mean across repetitions, $|\gamma_i(t)| < 1$ (since $d\theta_i/dt$ cannot be negative), $t \in [0, T_i]$, $\theta_i(0) = 0$, $\theta_i(T_i) = 1$, and T_i is the new trajectory length. Integration of this equation produces

$$\theta_i(t) = t + \phi_i(t) \tag{3}$$

where $\phi_i(t)$ is the phase variability function

$$\phi_i(t) = \int_0^t \gamma_i(u)du. \tag{4}$$

We also assume that there is some amplitude variability added to the clock's output, introduced by variations in the amplitude or shape of the lip movements. Then, each trajectory becomes

$$x_i(t) = p[\theta_i(t)] + \beta_i(t) \tag{5}$$

where $\beta_i(t)$ is a random function with zero mean across repetitions.

Combining Equations 5 and 3 we obtain the final model

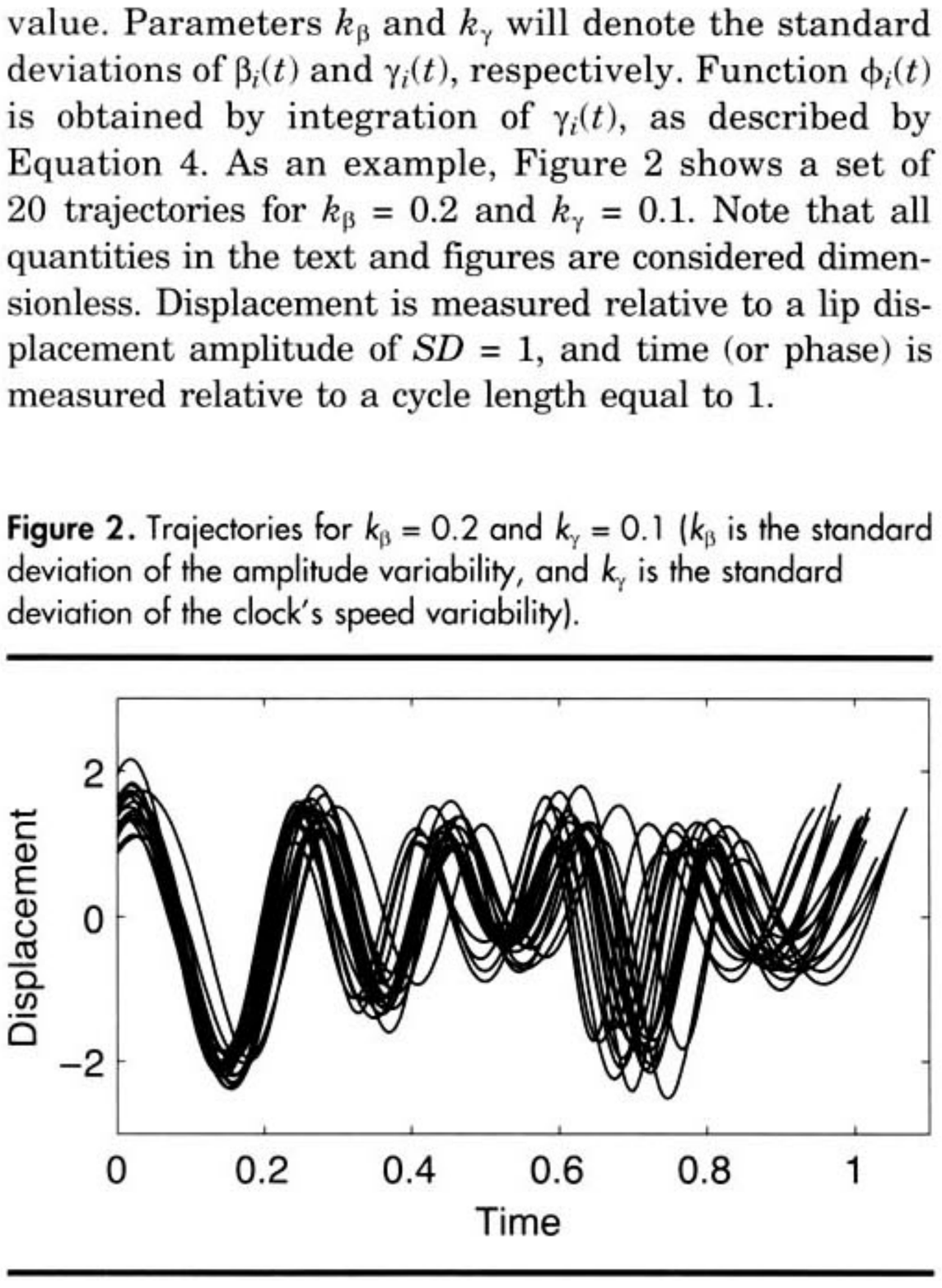
$$x_i(t) = p[t + \phi_i(t)] + \beta_i(t), \tag{6}$$

which will be used in the following sections to generate speech movement trajectories.

The above explanation has used a clock frequency of $\omega_p = 1$. Assuming that this value is in hertz units, then a lip trajectory pattern of a duration of 1 s would result, which approximately matches the duration of utterance "Buy Bobby a puppy" produced at normal speaking rate (Smith et al., 1995). The pattern is produced using a sampling interval of 1/200, which also matches the sampling frequency of 200 Hz used by Lucero et al. (1997) in their data collection.

Functions $\beta_i(t)$ and $\gamma_i(t)$ are produced by generating time series from a normal distribution with zero mean. The series are next smoothed by filtering them with a fifth-order low-pass Butterworth filter at 5-Hz cutoff. This value was selected by visual inspection of the results so as to obtain sets of trajectories that were visually similar to sets recorded experimentally (Lucero et al., 1997). It is also in agreement with results by Smith et al. (2000), which show that in such lip trajectories almost all of the energy (over 93%) is contained within that frequency range. Finally, the standard deviations of the filtered time series are set to a known value. Parameters k_β and k_γ will denote the standard deviations of $\beta_i(t)$ and $\gamma_i(t)$, respectively. Function $\phi_i(t)$ is obtained by integration of $\gamma_i(t)$, as described by Equation 4. As an example, Figure 2 shows a set of 20 trajectories for $k_\beta = 0.2$ and $k_\gamma = 0.1$. Note that all quantities in the text and figures are considered dimensionless. Displacement is measured relative to a lip displacement amplitude of $SD = 1$, and time (or phase) is measured relative to a cycle length equal to 1.

Figure 2. Trajectories for $k_\beta = 0.2$ and $k_\gamma = 0.1$ (k_β is the standard deviation of the amplitude variability, and k_γ is the standard deviation of the clock's speed variability).



The model above is proposed just as a simple mathematical means to produce sets of waveforms sharing a common pattern, and with known components of amplitude and phase variability. Although it reproduces lip trajectory waveforms collected experimentally, the model does not intend to provide a representation of the underlying physiology that produces the trajectories. Also, it does not say anything about how or at which point in the speech production process the variability is actually introduced in the planned pattern. Relations that might actually exist between phase, amplitude variability, and the trajectory shape are missed by the model. Its purpose is just to provide a simple means to test techniques for variability assessments, as will be done in the next section.

Extraction of the Pattern and Variability Components

The problem to consider here may be stated as follows: Given a set of N trajectories $x_i(t)$, $i = 1,\ldots,N$, described by the above equations, extract the pattern p and the variability introduced by functions β_i and ϕ_i. This problem is, in general, indeterminate; that is, there are infinite combinations of functions p, β_i, and ϕ_i that reproduce a given set of trajectories. However, a good approximation to its solution may be obtained, provided that the variability terms ϕ_i and β_i are small enough so that they do not produce a large distortion of pattern p in the trajectories x_i (i.e., provided that the trajectories x_i have the same general "shape" as p).

To solve this problem, a nonlinear normalization technique based on FDA may be applied. Details of the algorithms and their practical implementation may be found in the indicated references (e.g., Lucero et al., 1997; Ramsay & Li, 1998; Ramsay & Silverman, 1997). This technique first expands linearly all trajectories to a common length (linear normalization), which is set to a normalized value of 1. Next, a set of optimal transformations of time (warping functions) $h_i(t)$ is sought, such that the distance of each normalized trajectory $x_i^*(t) = x_i[h_i(t)]$ to their average $\bar{x}^*(t)$ is minimized while satisfying a roughness penalty constraint.

Ideally, one would like to determine warping functions satisfying

$$h_i^{-1}(t) = \theta_i(t) = t + \phi(t). \tag{7}$$

In this case, the normalized trajectories $x^*(t)$ reduce to

$$x^*(t) = p(t) + \beta_i^*(t) \tag{8}$$

where $\beta_i^*(t) = \beta_i[h_i(t)]$. Function $\beta_i^*(t)$ represents the amplitude deviation of each trajectory from the pattern, at each instant of time. The pattern may be computed as the average of the normalized trajectories; that is, $x^*(t) = p(t)$, assuming that $\beta_i^*(t)$ also has zero mean across repetitions, and $\beta_i^*(t)$ simply becomes the difference of each normalized trajectory to the average.

Letting $\phi_i^*(t) = \phi_i[h_i(t)]$, we have from Equation 7

$$\phi_i^*(t) = t - h_i(t). \tag{9}$$

This function represents the time shift of each point in the pattern p to the position of the same point in trajectory i.

The normalized variability functions $\beta_i^*(t)$ and $\phi_i^*(t)$ may be used to visualize the distribution of amplitude and phase variability, respectively, along the trajectory pattern, and compute suitable indices, as will be done in the next section. If required, a frequency variability function may be also computed as $\gamma_i^*(t) = d\phi_i^*(t)/dt$, which represents variability in the speed of the clock (or speaking rate) that produces the trajectories.

In the simple case that there is no amplitude variability, that is, $\beta_i(t) \equiv 0$, and the measure of distance of the normalized trajectories to their average for the nonlinear normalization algorithm is the mean square error

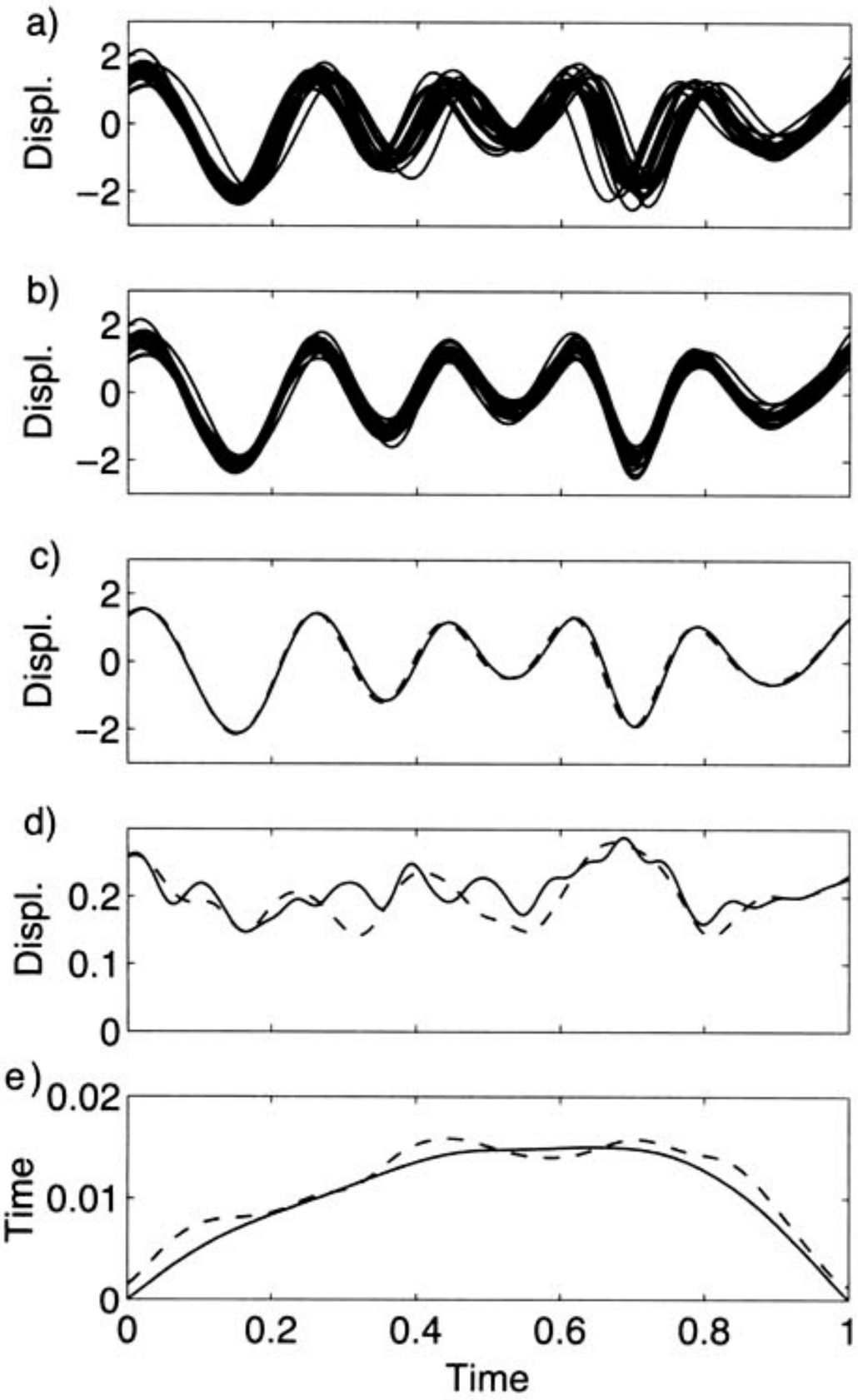
$$D = \sum_{i=0}^{N} \int_0^1 \left[x_i^*(t) - \bar{x}^*(t)\right]^2 dt, \tag{10}$$

then the warping functions in Equation 7 are a minimizer of D, producing $D = 0$, because $x_i^*(t) = \bar{x}^*(t) = p(t)$. It may be also shown that any other warping function such that $\delta(t) = h_i(t) + \phi_i[h_i(t)]$ for an arbitrary function $\delta(t)$ will also produce $D = 0$ (Wang & Gasser, 1997).

In practice, several factors complicate the computation of the above ideal warping functions. In general, $\beta_i(t)$ is not identically zero. Also, the measure of closeness used by the nonlinear normalization algorithm includes a roughness penalty term, to avoid excessive distortion of the trajectories (Lucero et al., 1997; Smith et al., 1995). An additional complication is that the algorithm might get trapped in local minima and produce a warping function that is not optimal. Thus, we may question how close the computed warping functions are to the desired functions $h_i(t) = \theta_i^{-1}(t)$.

We consider this question through a numerical example. Figure 3 shows results when the FDA nonlinear normalization is applied to the set of wavelets in Figure 2. We can see (Figure 3c) that the extracted pattern is very close to the actual pattern. The extraction of amplitude and phase variability functions is evaluated through their standard deviations (Figures 3d and 3e,

Figure 3. Nonlinear normalization results for wavelets in Figure 2. (a) Linearly normalized trajectories, (b) nonlinearly normalized trajectories, (c) pattern, (d) standard deviation of amplitude variability functions across trajectories, and (e) standard deviation of phase variability functions across trajectories. Curves in full line correspond to computed functions, and curves in broken line correspond to actual functions.



respectively). These curves were produced by computing the standard deviations of the variability functions across trajectories at each sample point in time, and they represent the distribution of variability along the pattern length. In the case of the phase variability, note that the initial and ending values are zero. This is a consequence of the linear normalization that is applied at the start of the FDA algorithm (Figure 3a). Since all the linearly normalized trajectories start and end at the same instant of time, then the phase variability at those instants must be zero.

The computed phase variability provides a good approximation to the actual curve. In the case of the amplitude variability, we can note differences of shape details between the computed and actual curves. However, the general shape and mean level of both curves are close. The mean values of the standard deviation curves for the actual amplitude and phase variability are 0.2002 and 0.0113, respectively. The means of the computed standard deviations were 0.1879 and 0.0110, respectively, with errors of 6.1% and 2.6%.

Indices of Variability

After the variability functions have been extracted, various indices may be defined and computed for quantitative assessment. For example, we may use the same measures above: an index of amplitude variability (IAV), defined as the mean of the standard deviation of the amplitude variability functions $\beta_i^*(t)$, and an index of phase variability (IPV), defined as the mean of the standard deviation of the phase variability functions $\phi_i^*(t)$.

Figures 4 and 5 show the computed and actual values of IAV and IPV for a range of values of k_β and k_γ, respectively. For comparison with previous approaches for assessing variability (e.g., Smith et al., 1995, 2000; Ward & Arnfield, 2001), the figures also show the IAV computed as the mean of the standard deviation of the unnormalized trajectories (to compute the standard deviation, the trajectories are first zero-padded to a common length) and of the linearly and piecewise linearly normalized trajectories.

Linear normalization was used by Smith et al. (1995, 2000) to compute an index of spatiotemporal stability (STI). The STI is computed by using Fourier interpolation to linearly normalize the trajectories and they taking the sum of the standard deviations at 2% intervals of the total trajectory length. It may be easily shown that the STI is proportional to the linearly normalized IAV defined above (small differences may appear due to a waveform smoothing introduced by the Fourier interpolation; see a discussion in Lucero et al., 1997).

Piecewise linear normalization was proposed recently by Ward and Arnfield (2001). In this case, specific events of the trajectories, such as the maxima of the curves, are first selected and aligned. Next, linear normalization is applied between those prealigned events. Finally, the sum of the standard deviations at 2% intervals of the total trajectory is computed as a variability index. Here, the same piecewise technique was applied to the trajectories, but the index was computed as defined above, to allow comparison of the results.

Figure 4 shows variability indices when varying the amplitude variability (k_β), while the clock speed variability is held constant at $k_\gamma = 0.1$. Figure 4a shows

Figure 4. Indices of variability versus k_β, for $k_\gamma = 0.1$ (constant). (a) Index of amplitude variability and (b) index of phase variability. Circles: actual values; triangles: computed values with unnormalized trajectories; crosses: computed values with linear normalization; stars: computed values with piecewise linearly normalized trajectories; squares: computed values with nonlinearly normalized trajectories.

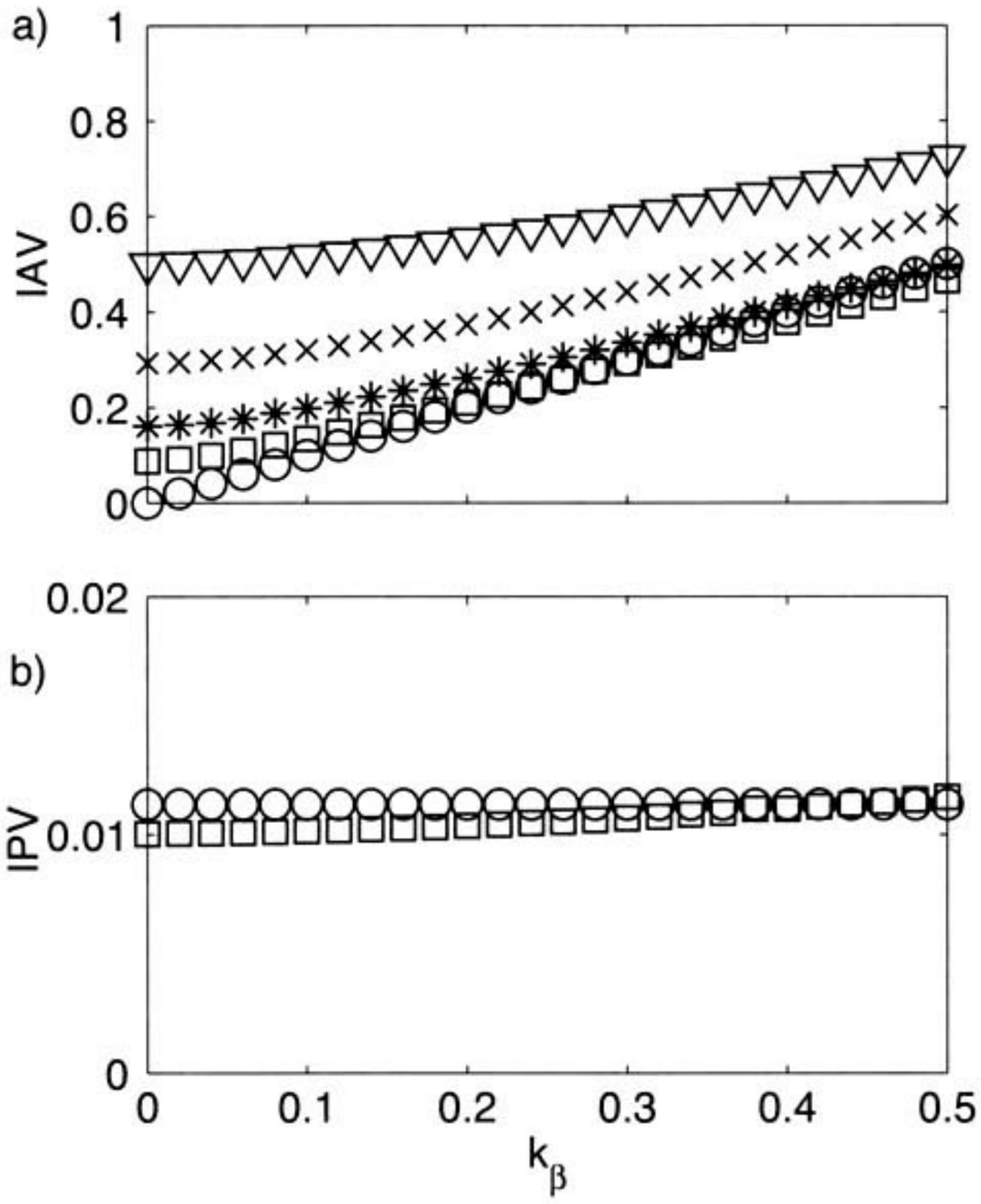
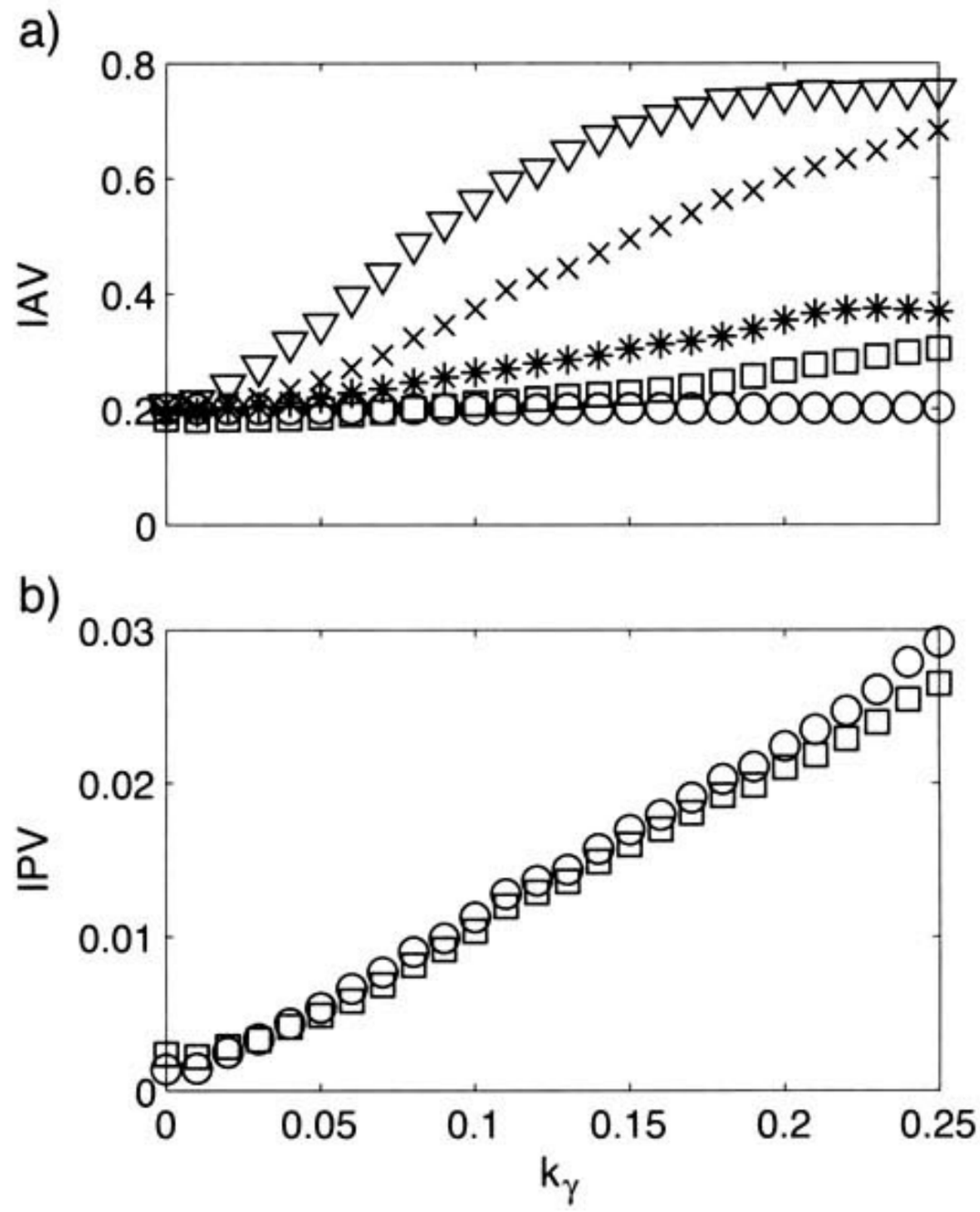


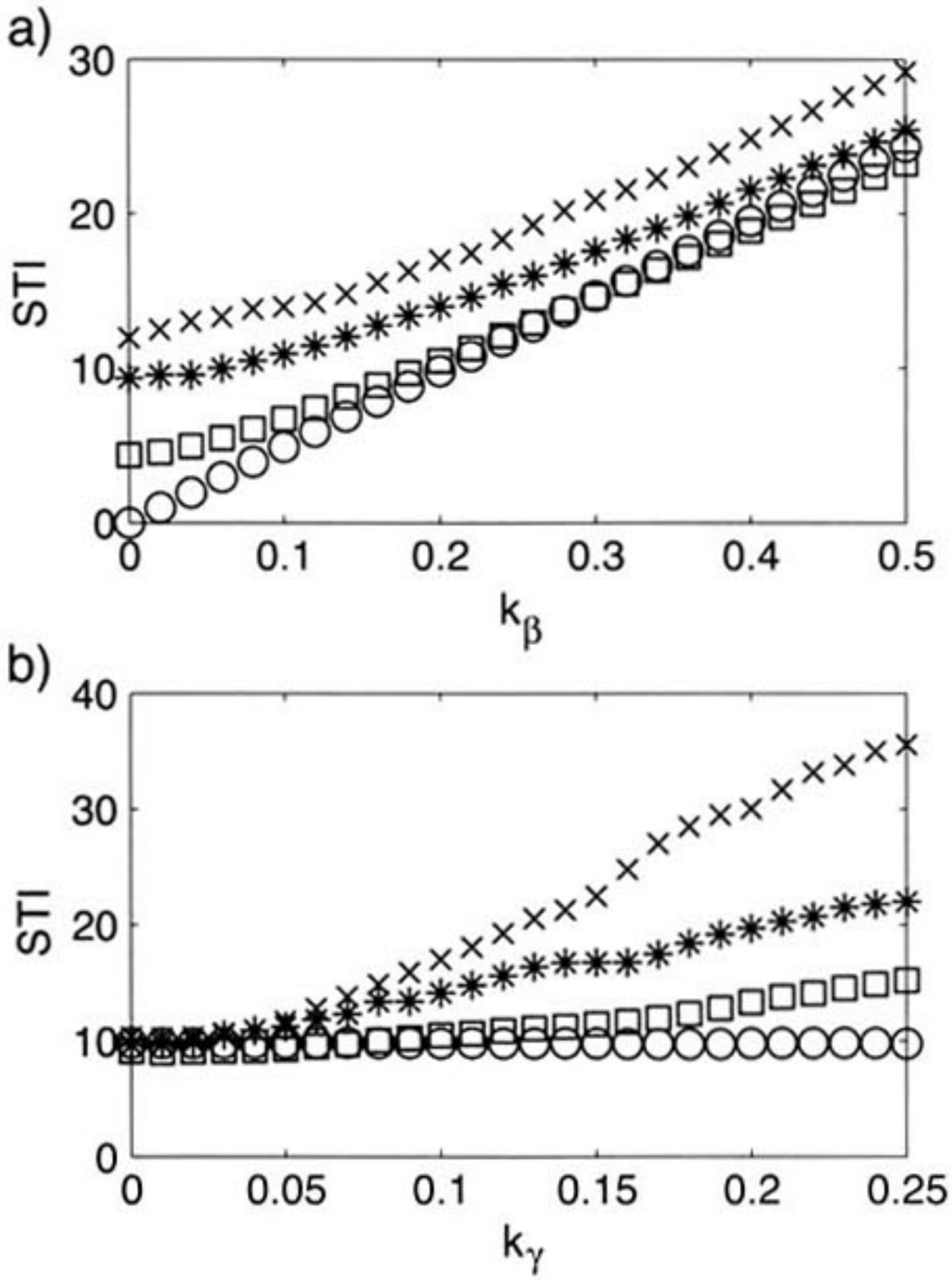
Figure 5. Indices of variability versus k_γ, for $k_\beta = 0.2$ (constant). (a) Index of amplitude variability and (b) index of phase variability. Circles: actual values; triangles: computed values with unnormalized trajectories; crosses: computed values with linear normalization; stars: computed values with piecewise linearly normalized trajectories; squares: computed values with nonlinearly normalized trajectories.



that all indices increase with k_β, as expected. The nonlinearly normalized IAV provides in general the best approximation to the actual IAV value. The linearly normalized IAV overestimates the amplitude variability, because this measure is contaminated by the phase variability of the trajectories. The same effect occurs with the unnormalized IAV, which provides the worst approximation. The piecewise linear IAV falls between the linear and nonlinear indices, since it may be considered a hybrid of both approaches. However, it is much closer to the nonlinear index and provides good approximations to the actual variability at large values. As the amplitude variability grows large and the phase variability becomes negligible in comparison, all indices tend to converge. In the case of a zero amplitude variability, even the unnormalized IAV has a positive value, as a consequence of errors introduced by the phase variability of the set. Figure 4b shows that the nonlinearly normalized IPV provides a good approximation to the constant phase variability of the set. At low amplitude variabilities, the IPV underestimates the phase variability, because part of it is measured as amplitude variability by the nonlinear IAV. For very large amplitude variabilities, the opposite effect occurs.

Figure 5 shows similar results when varying the clock speed variability (k_γ) and consequently the phase variability, while the amplitude variability is held constant at $k_\beta = 0.2$. It shows that the nonlinearly normalized IAV provides a good approximation to an almost constant amplitude variability. The linearly normalized IAV again overestimates the variability and increases almost linearly with k_γ, showing that this index measures a combination of both amplitude and phase variability. The same increase occurs with the piecewise linear and the unnormalized IAV, although these two indices seem to reach a saturation level at large variability. Some of the indices also seem to converge at large values of k_γ; however, note that they all diverge from the actual IAV value and become increasingly poorer. Again, the piecewise linear IAV falls between the nonlinear and linear indices.

Figure 6. (a) STI versus k_β, for $k_\gamma = 0.1$ (constant) and (b) STI versus k_γ, for $k_\beta = 0.2$ (constant). Circles: actual values; crosses: computed values with linear normalization; stars: computed values with piecewise linearly normalized trajectories; squares: computed values with nonlinearly normalized trajectories.



Finally, and for better comparison with the STI measure of previous works, this index was computed for the sets of synthetic trajectories in four different ways:

1. Linear normalization, following the technique described by Smith et al. (1995). First, the segment between the first and last negative peaks of lip velocity (i.e., the peaks of the first and last mouth-opening movements) was extracted from each record. Next, all segments were linearly normalized to 1,000 sample points using Fourier interpolation with 10 harmonics. Finally, the STI was computed as the sum of the standard deviation across records at 2% intervals of the record length.
2. Piecewise linear normalization, following the technique described by Ward and Arnfield (2001). In this case, segments were first extracted from each record as in the previous way. Next, the maxima of the lip displacement curves were identified and aligned, while linearly normalizing the intervals between maxima to produce a common total length of 1,000 points. The linear normalization was performed using cubic spline interpolation. Finally, the STI was computed as in the previous way.
3. In this case, the extracted segments were nonlinearly normalized to a common length of 1,000 points using the FDA technique, and the STI was next computed as in the first way.
4. The actual value of the STI was computed from the amplitude variability functions $\beta_i^*(t)$.

Figures 6a and 6b show the computed STI results for the same variability levels as in Figures 4 and 5, respectively. The STI values approximately follow the same behavior as the indices of the previous figures, except for a change of vertical scale and small differences caused by the initial segment extraction procedure. In general, we see that the nonlinearly normalized STI provides a better estimation of the amplitude variability of the set of trajectories than the linearly normalized STI, which is affected also by the phase variability of the set. The piecewise linearly normalized STI values fall between the other two indices.

Discussion

The results have shown that the FDA nonlinear normalization technique permits us to extract the pattern and variability of a set of speech movement trajectories with good approximation, when those trajectories are modeled as in Equation 6. Amplitude and phase variability are extracted as separate functions, which may be used to visualize the distribution of variability along the trajectory pattern. The variability of the set may be also assessed through the proposed IAV and IPV indices, which measure the actual amplitude and phase variability with good accuracy.

One must be careful when looking at details of the distribution of variability along the trajectory pattern. The plot of amplitude variability in Figure 3d shows that, although the general level and waveshape of the variability distribution are recovered by the nonlinear normalization, there are still differences in the details of that shape. This might be an important issue to consider with care when looking for regions that are more and less variable within an utterance.

Indices of variability based on linear normalization, such as the STI (Smith et al., 1995), measure a combination of amplitude and phase variability. Thus, the same measure may be produced by different levels of amplitude and phase variability. For example, the combinations $k_\gamma = 0.1$, $k_\beta = 0.5$ (see Figures 4a and 6a) and $k_\gamma = 0.2$, $k_\beta = 0.2$ (see Figures 5a and 6b) produce the same linearly normalized IAV around 0.6, and

linearly normalized STI around 30, although they have very different components of variability. Also, two sets of trajectories with the same level of amplitude and phase variability but different shape (pattern) may produce different variability indices (Lucero & Koenig, 2000).

The piecewise linearly normalized technique produces results that are between the linear and the nonlinear normalization techniques. A disadvantage of this technique is that it requires the user to select specific events for the initial alignment, whereas such selection is not required in either the linear or the nonlinear techniques. A proper selection of events may be unclear, for example, in the case of multidimensional records of simultaneous articulatory movements (e.g., Koenig et al., 2003; Lucero & Löfqvist, 2003). Also, the resultant transformations of the time scale (warping functions) are piecewise linear and consequently not smooth. Smoothness of the results might be a desired characteristic, for example, for further processing involving derivatives of the computed functions.

The model used to produce the speech movement trajectories assumes that amplitude and phase variability are added to the trajectory pattern as independent components. However, other models for introducing variability to speech trajectories might be used. For example, variability in muscle forces acting on the lips will translate into both amplitude and phase variability of the lip trajectories, in a way that depends on their biomechanical characteristics (Lucero, 2002). To analyze such a case, the model proposed here to generate speech trajectories is not adequate, because amplitude and phase variability will not be independent, and a different procedure based on dynamical models should be used. Such models might include theoretical models of the lip dynamical structure (Gomi, Nozoe, Dang, & Honda, 2003; Muller, Milenkovic, & MacLeod, 1985), empirical models based on descriptions of the trajectories using differential equations (Lucero, 2002; Ramsay & Silverman, 1997), full models of facial biomechanics (Lucero & Munhall, 1999), statistical and neural network models of speech movement production (Kuratate, Munhall, Rubin, Vatikiotis-Bateson, & Yehia, 1999; Vatikiotis-Bateson & Yehia, 1996), and others.

Similarly, other measures of variability are possible and deserve further exploration. For example, it has been proposed that direct computation of variability at the forcing level might permit a better assessment of variability than measures at the displacement level (Lucero, 2002). Such computation may be performed by using principal differential analysis (Ramsay & Silverman, 1997), which is another FDA technique based on fitting a differential equation to the set of trajectories, to extract their forcing functions.

Acknowledgments

This work was supported by the Conselho Nacional de Desenvolvimento Cientifico e Tecnologico (Brazil). The results reported here were presented in partial form at the 5th Seminar on Speech Production, in Kloster Seeon, Germany, May 2000. I am grateful to Dr. Kevin G. Munhall and Dr. Vincent L. Gracco for their interest and useful discussions on this research.

References

Gomi, H., Nozoe, J., Dang, J., & Honda, K. (2003). Physiologically based lip model for generating speech articulation. In S. Palethorpe & M. Tabain (Eds.), *Proceedings of the 6th International Seminar on Speech Production* (pp. 79–84). Sydney, Australia: Macquarie University.

Koenig, L. L., & Lucero, J. C. (2002). Oral-laryngeal control patterns for fricatives in 5-year olds and adults. In J. H. L. Hansen & B. Pellom (Eds.), *Proceedings of the 7th International Conference on Spoken Language Processing* (pp. 49–52). Denver, CO: Causal Productions.

Koenig, L. L., Lucero, J. C., & Löfqvist, A. (2003). Studying articulatory variability using functional data analysis. In M. J. Solé, D. Recasens, & J. Romero (Eds.), *Proceedings of the 15th International Conference on Phonetic Science* (pp. 269–272). Barcelona, Spain: Causal Productions.

Kuratate, T., Munhall, K., Rubin, P., Vatikiotis-Bateson, E., & Yehia, H. (1999). Audio-visual synthesis of talking faces from speech production correlates. *Proceedings of the 6th European Conference on Speech Communication and Technology (EUROSPEECH'99), 3,* 1279–1282.

Lucero, J. C. (2002). Identifying a differential equation for lip motion signals. *Medical Physics and Engineering, 24,* 521–528.

Lucero, J. C., & Koenig, L. L. (2000). Time normalization of voice signals using functional data analysis. *Journal of the Acoustical Society of America, 108,* 1408–1420.

Lucero, J. C., & Löfqvist, A. (2003). Functional data analysis of articulatory variability in vcv sequences. In S. Palethorpe & M. Tabain (Eds.), *Proceedings of the 6th International Seminar on Speech Production* (pp. 156–160). Sydney, Australia: Macquarie University.

Lucero, J. C., & Munhall, K. G. (1999). A model of facial biomechanics for speech production. *Journal of the Acoustical Society of America, 106,* 2834–2842.

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

Muller, E. M., Milenkovic, P. H., & MacLeod, G. E. (1985). Perioral tissue mechanics during speech production. In C. DeLisi & J. Eisendfeld (Eds.), *Mathematics and*

computers in biomedical applications (pp. 363–371). Amsterdam: Elsevier.

Neilson, P. D., & O'Dwyer, N. J. (1984). Reproducibility and variability of speech muscle-activity in athetoid dysarthria of cerebral-palsy. *Journal of Speech and Hearing Research, 27,* 502–517.

Qi, Y. (1992). Time normalization in voice analysis. *Journal of the Acoustical Society of America, 92,* 2569–2576.

Qi, Y., Weinberg, B., Bi, N., & Hess, W. J. (1995). Minimizing the effect of period determination on the computation of amplitude perturbation in voice. *Journal of Acoustical Society of America, 97,* 2525–2532.

Ramsay, J. O., & Li, X. (1998). Curve registration. *Journal of the Royal Statistical Society, Series B, 60,* 351–363.

Ramsay, J. O., & Silverman, B. W. (1997). *Functional data analysis*. New York: Springer-Verlag.

Smith, A., Goffman, L., Zelaznik, H. N., Ying, G., & McGillem, C. (1995). Spatiotemporal stability and patterning of speech movement sequences. *Experimental Brain Research, 104,* 439–501.

Smith, A., Johnson, M., McGillem, C., & Goffman, L. (2000). On the assessment of stability and patterning of speech movements. *Journal of Speech, Language, and Hearing Research, 43,* 277–286.

Strick, H., & Boves, L. (1991). A dynamic programming algorithm for time-aligning and averaging physiological related to speech. *Journal of Phonetics, 19,* 367–378.

Titze, I. R., & Liang, H. (1993). Comparison of f_0 extraction methods for high-precision voice perturbation measurements. *Journal of Speech, Language, and Hearing Research, 36,* 1120–1133.

Vatikiotis-Bateson, E., & Yehia, H. C. (1996). Physiological modeling of facial motion during speech. *Transactions of the Technical Committee on Psychological and Physiological Acoustics, H-96, 65,* 1–8.

Wang, K., & Gasser, T. (1997). Alignment of curves by dynamic time warping. *Annals of Statistics, 25,* 1251–1276.

Ward, D., & Arnfield, S. (2001). Linear and nonlinear analysis of the stability of gestural organization of speech movement sequences. *Journal of Speech, Language, and Hearing Research, 44,* 108–117.

Yumoto, E., Gould, W. J., & Baer, T. (1982). Harmonics-to-noise ratio as an index of degree of hoarseness. *Journal of the Acoustical Society of America, 71,* 1544–1550.

Received May 10, 2004

Accepted August 17, 2004

DOI: 10.1044/1092-4388(2005/023)

Contact author: Jorge C. Lucero, Department of Mathematics, University of Brasilia, Brasilia, DF 70910-900, Brazil. E-mail: lucero@mat.unb.br