Time normalization of voice signals using functional data analysis

Jorge C. Lucero[a]
Department of Mathematics, University of Brasilia, Brasilia DF 70910-900, Brazil

Laura L. Koenig[b]
Haskins Laboratories, 270 Crown Street, New Haven, Connecticut 06511 and Long Island University, Brooklyn, New York 11201-8423

(Received 6 October 1999; accepted for publication 28 March 2000)

The harmonics-to-noise ratio (HNR) has been used to quantify the waveform irregularity of voice signals [Yumoto *et al.*, J. Acoust. Soc. Am. **71**, 1544–1550 (1982)]. This measure assumes that the signal consists of two components: a harmonic component, which is the common pattern that repeats from cycle-to-cycle, and an additive noise component, which produces the cycle-to-cycle irregularity. It has been shown [J. Qi, J. Acoust. Soc. Am. **92**, 2569–2576 (1992)] that a valid computation of the HNR requires a nonlinear time normalization of the cycle wavelets to remove phase differences between them. This paper shows the application of functional data analysis to perform an optimal nonlinear normalization and compute the HNR of voice signals. Results obtained for the same signals using zero-padding, linear normalization, and dynamic programming algorithms are presented for comparison. Functional data analysis offers certain advantages over other approaches: it preserves meaningful features of signal shape, produces differentiable results, and allows flexibility in selecting the optimization criteria for the wavelet alignment. An extension of the technique for the time normalization of simultaneous voice signals (such as acoustic, EGG, and airflow signals) is also shown. The general purpose of this article is to illustrate the potential of functional data analysis as a powerful analytical tool for studying aspects of the voice production process. © *2000 Acoustical Society of America.* [S0001-4966(00)00310-6]

PACS numbers: 43.70.Aj, 43.70.Dn, 43.70.Gr, 43.72.Lc [AL]

I. INTRODUCTION

This paper deals with the problem of quantifying the irregularity in the waveform of a voice signal. It has long been known that measures of irregularity in the time and/or amplitude domain may differentiate normal from abnormal voice qualities, with the pathological samples showing more extreme measures of irregularity than the normal samples (Lieberman, 1962; Titze, 1994a). Thus accurate measures of waveform irregularity could be used as a noninvasive technique for voice evaluation and diagnosis.

As pointed out by Qi (1992), computing such measures of waveform irregularity presents the difficulty that an infinite amount of information is involved, in contrast to, e.g., measures of fundamental frequency irregularity (jitter) which deal with a single parameter. The simplest approach is to compute the variability on the maximum amplitude of each period (wavelet) of the signal. However, this measure has limitations since it misses information at other points of the wavelets. It is easy to see that wavelets of different shapes but the same maximum amplitude would produce a zero measure of irregularity by such an approach.

As an improved measure, the harmonics-to-noise ratio (HNR; Yumoto *et al.*, 1982) was proposed, in which the whole wavelet is used in the computation. The HNR assumes that the signal consists of two components: a harmonic component which is the periodic pattern that repeats through all the wavelets, and an additive noise component which produces wavelet irregularity. In the cited work, the harmonic component was computed as the average of the wavelets, and the noise component as the difference of the wavelets to their average. Since the wavelets have different lengths due to jitter, they were normalized in time by zero padding (i.e., filling with zeroes) each wavelet to the longest period, so that they could be compared on a point-by-point basis.

Qi (1992) showed the limitations of the zero-padding normalization: since the wavelets differ in length, a large portion of the computed noise will be caused by the length irregularity. Thus voices with high values of jitter will necessarily produce low values of the HNR, so that the HNR in such cases does not provide an accurate indication of general waveform irregularity. A first solution to this problem would be a linear expansion or compression of all wavelets to a common length. However, phase differences between wavelets would remain, which would also contaminate the computed HNR. To illustrate this problem, a simple case of two wavelets is shown in Fig. 1. In each plot, the broken line is the computed average. In the case of zero-padding normalization (Fig. 1, top), the average clearly does not resemble either of the wavelets. It also has a point of discontinuity at the start of the zero-padding region. With linear normalization (Fig. 1, middle), a better continuous average is obtained, although its shape is still different from those of the wavelets. To obtain a better average, phase differences between the wavelets should be removed. A more accurate computa-

[a]Electronic mail: lucero@mat.unb.br
[b]Electronic mail: koenig@haskins.yale.edu

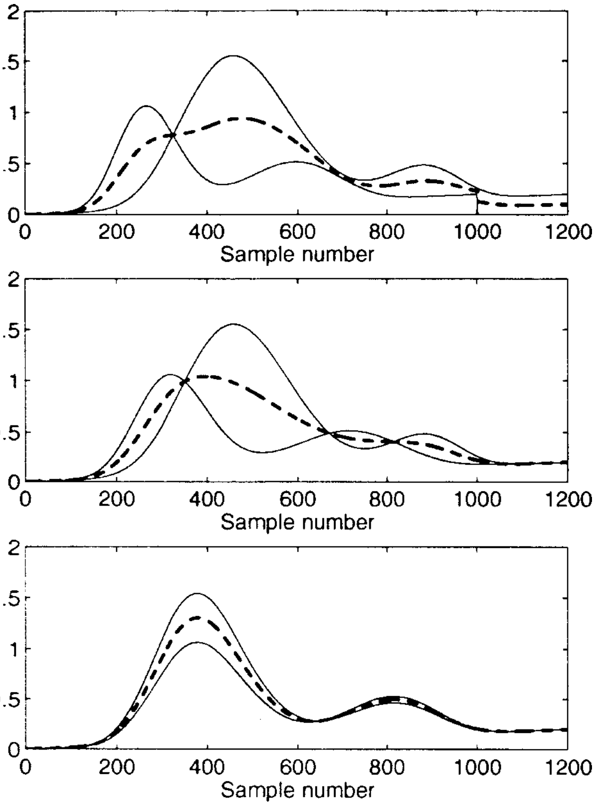

FIG. 1. Methods of temporal normalization applied to the extraction of the average (dashed line) of two wavelets. Top: zero padding. Middle: linear normalization. Bottom: nonlinear normalization.

tion of the wavelet average and the HNR requires a nonlinear expansion or compression of the wavelets in time, so that their shapes become aligned (Fig. 1, bottom). Only in this case may the average be considered as the common pattern of the wavelets. To accomplish an optimal wavelet alignment, Qi (1992) applied a dynamic programming algorithm. In later works (Qi *et al.*, 1995; Qi and Hillman, 1997), unconstrained dynamic programming and zero phase transformation were used for the alignment. The zero phase transformation simply removes all phase-related information from the wavelets prior to computation of the HNR; however, this approach produced in general poorer measures of waveform irregularity than nonlinear normalization, according to tests using synthetic signals (Qi *et al.*, 1995).

A similar issue has been recently discussed in the case of speech movement signals (Lucero *et al.*, 1997). In that work, three techniques for extracting the average of a set of speech wavelets were considered, namely: un-normalized averaging (equivalent to zero padding, Fig. 1, top), linearly normalized averaging (as in Fig. 1, middle), and nonlinearly normalized averaging (as in Fig. 1, bottom). To achieve the nonlinearly normalized average, a new algorithm based on functional data analysis (FDA; Ramsay, 1998; Ramsay *et al.*, 1996; Ramsay and Li, 1998; Ramsay and Silverman, 1997) was introduced. It was argued that this algorithm has advantages over previous dynamic programming because the results are smooth and differentiable (thus allowing for further processing), it does not require users to select one of the wavelets as a reference or template for the alignment, and different optimization criteria may be adopted according to the application.

This work will show the application of the FDA nonlinear normalization technique to extract averages and compute the HNR of voice signals. Results obtained for the same signals using zero-padding, linear normalization, and dynamic programming algorithms will also be presented for comparison. Our general purpose is to illustrate the potential of FDA as a powerful analytical tool for studying aspects of the voice production process.

II. MEASUREMENT OF VOICE SIGNAL IRREGULARITY USING FDA

A. Nonlinear time normalization

FDA has emerged in recent years as a set of analytical tools to explore patterns and variability in sets of data that may be regarded as functional observations (Ramsay and Silverman, 1997). The term functional here means that, although the data may be observed and recorded discretely, they may be described by some function of time. A single functional observation z consists of a finite set of pairs (t_j, z_j), where z_j is the measured jth sample of z at time t_j. In FDA, the existence of an underlying function $y(t)$ is postulated, such that

$$z_j = y(t_j) + \epsilon_j, \tag{1}$$

where ϵ_j represents an observational error or noise term. A variety of analytical tools may be applied to extract the main characteristics of the functional data set. Such tools may require evaluating such a function $y(t)$ at any particular instant of time, and all its derivatives that exist at such an instant. Two approaches may then be followed: (1) extracting $y(t)$ from the raw data by filtering out the noise (i.e., by smoothing the data), or (2) leaving the noise in the data and requiring smoothness of the results of the analysis. In the present case, smoothing the raw data would eliminate or attenuate the same irregularity we want to assess, so the second approach will be followed. We will align the raw voice wavelets by requiring a smooth expansion or compression of the time scale. We describe briefly the FDA algorithm for nonlinear time normalization. For more details, we refer the reader to the cited references (Lucero *et al.*, 1997; Ramsay, 1998; Ramsay and Li, 1998; Ramsay and Silverman, 1997).

Let us denote the set of wavelets to normalize as $x_i(t)$, where $i = 1,...,N$, and N is the number of wavelets. For simplicity, let us assume that all the wavelets have the same length, from $t = 0$ to $t = 1$. For each wavelet, a strictly increasing and smooth transformation of time $h_i(t)$ (warping function) is determined, such that each normalized wavelet

$$x_i^*(t) = x_i[h_i(t)] \tag{2}$$

is close in some measure to their average

$$\overline{x^*}(t) = \frac{1}{N} \sum_{i=1}^{N} x_i^*(t). \tag{3}$$

Such a transformation is defined as

$$h_i(t)=A\int_0^t e^{\int_0^u w_i(v)dv}\,du, \tag{4}$$

where $w_i(t)$ is the relative curvature of $h_i(t)$ (to be determined optimally), v is an integration variable, and coefficient A is selected so that $h_i(t)=1$. Given any function $w_i(t)$ such that the integrals in Eq. (4) exist, this equation will produce a strictly increasing and twice differentiable function $h_i(t)$.

Different measures may be used to evaluate the closeness of the normalized records to their average, according to the particular application. Here, the measure

$$F(x_i,w_i,\alpha,\lambda)=\int_0^1 \alpha(t)[x_i^*(t)-\overline{x}^*(t)]^2\,dt+\lambda\int_0^1 w_i^2(t)dt \tag{5}$$

is adopted, where $\alpha(t)$ is a weighting function and λ is a positive constant. The first integral is the classic squared error measure used in dynamic programming algorithms (Qi, 1992; Qi *et al.*, 1995; Qi and Hillman, 1997). The weighting function $\alpha(t)$ may be used to emphasize alignment in particular regions of the wavelets [by setting a larger value of $\alpha(t)$ at those regions]. The second integral incorporates a penalty for the roughness of the warping function, controlled by parameter λ (the larger the value of λ, the smaller the curvature of h_i).

Hence, the problem consists of estimating the curvature functions $w_i(t)$ in Eq. (4) that will minimize the total measure (cost function)

$$C(x_1,\ldots,x_N,w_1,\ldots,w_N,\alpha,\lambda)=\sum_{i=1}^{N} F(x_i,w_i,\alpha,\lambda). \tag{6}$$

This minimization problem may be solved by using an expansion of $\int_0^1 w_i(t)dt$ into a basis of B-spline functions, as described by Ramsay and Silverman (1997) and Lucero *et al.* (1997).

The algorithm assumes that all the wavelets have the same length from 0 to 1. It is possible to modify it to accommodate wavelets of different lengths and time spans. However, it is computationally much simpler to interpolate all wavelets to a common length and attribute to this length an artificial [0, 1] time span before applying the nonlinear normalization. The results should be the same in either case.

After the wavelets have been optimally aligned in time, we extract the normalized average $\overline{x}^*(t)$ and the normalized noise components $x_i^*(t)-\overline{x}^*(t)$, $i=1,\ldots,N$. We also compute the expression $\Delta h_i(t)=h_i(t)-t$, $i=1,\ldots,N$, which represents the amount of nonlinear warping or phase shift for each wavelet [if no warping is required, then $h_i(t)=t$ and $\Delta h_i(t)=0$]. The irregularity of the set of wavelets may then be seen in the above functions or related ones. For example, one may compute the standard deviation of both $x_i^*(t)$ and $h_i(t)$ across the N wavelets, to visualize how the waveform irregularity and phase irregularity are distributed along the wavelet period [0, 1].

B. Extension to simultaneous signals

The above algorithm may be easily extended for simultaneous normalization of sets of signals. When various signals are recorded simultaneously (e.g., acoustic, EGG, oral airflow, and other voice signals), it might be desirable also to normalize them simultaneously, to keep their synchrony in time. Also, simultaneous normalization may be applied to reveal phase relations between the signal sets.

For this case, instead of scalar-valued wavelets, one may consider vector-valued wavelets such as

$$\mathbf{x}_i(t)=[\mathrm{Acoustics}_i(t),\mathrm{EGG}_i(t),\mathrm{Airflow}_i(t),\ldots]^T. \tag{7}$$

The warping functions are still scalar functions, which simultaneously align all the components of the wavelets.

The cost function has now the general expression

$$F(\mathbf{x}_i,w_i,\mathbf{A},\lambda)=\int_0^1[\mathbf{x}_i^*(t)-\overline{\mathbf{x}}^*(t)]^T\mathbf{A}(t)[\mathbf{x}_i^*(t)-\overline{\mathbf{x}}^*(t)]dt + \lambda\int_0^1 w_i^2(t)dt, \tag{8}$$

where $\mathbf{A}(t)$ is a matrix of weight functions.

C. Indices of irregularity

We consider here two indices of irregularity. The HNR (Yumoto *et al.*, 1982),

$$\mathrm{HNR}=\frac{N\int_0^1\overline{x}^{*2}(t)dt}{\Sigma_{i=1}^N\int_0^1[x_i^*(t)-x^*(t)]^2\,dt} \tag{9}$$

and an index of nonlinear warping (INW), defined as the mean of the root-mean-squared values of warping functions $\Delta h_i(t)$:

$$\mathrm{INW}=\frac{1}{N}\sum_{i=1}^{N}\sqrt{\int_0^1 \Delta h_i(t)dt}. \tag{10}$$

III. EXAMPLES WITH SYNTHETIC SIGNALS

A. Signals

We first applied the above techniques to synthetic signals, in order to determine the relative accuracy of the different approaches. To allow comparison with previous work, we synthesized signals following the equations given by Titze and Liang (1993).

We set the instantaneous frequency of the signals to

$$f_c(t)=f_0(1+K_f\sin(2\pi f_0 t/10)), \tag{11}$$

where f_0 is the center frequency, and K_f is a parameter for frequency variability. For simplicity, we adopted a sinusoidal frequency modulation instead of a random modulation. The instantaneous phase is

$$\theta(t)=\int_0^t 2\pi f_c(t)dt. \tag{12}$$

The signal is then synthesized as

$$x(t)=y[\theta(t)]+K_a\,\mathrm{Randn}(t), \tag{13}$$

where $y(\theta)$ is a periodic function, K_a is a parameter for amplitude variability of the signal noise, and $\mathrm{Randn}(t)$ is a function that generates a random value from a normal distri-

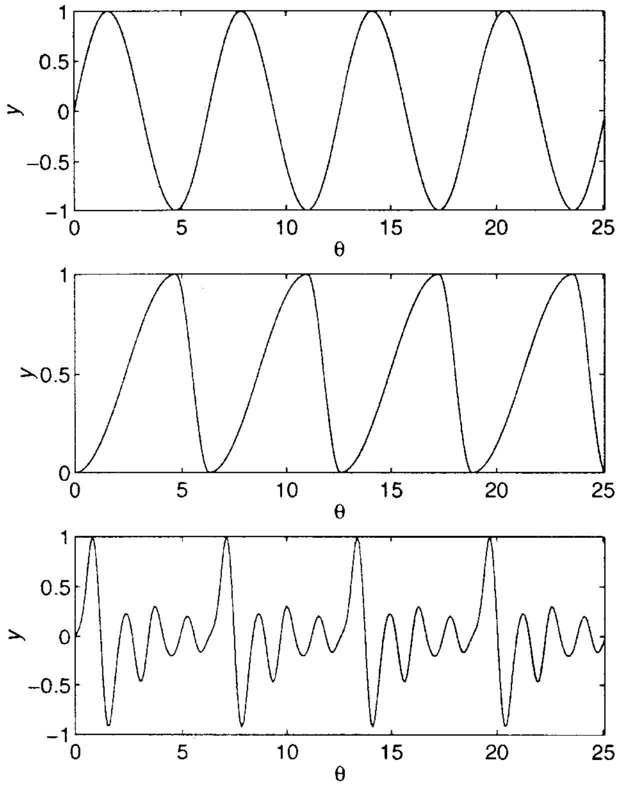

FIG. 2. Synthetic signals used to test the algorithms. Top: sine wave. Middle: EGG analog. Bottom: mouth pressure analog.

bution. We considered various functions for $y(\theta)$: (a) plain sine wave (Fig. 2, top)

$$y(\theta)=\cos(\theta), \tag{14}$$

(b) EGG analog (Fig. 2, middle), using the expression (Titze and Liang, 1993)

$$y(\theta)=\begin{cases} 0.5[1-\cos(2\theta)], & 0\leqslant\theta<\pi/2 \\ 0.5[1-\cos(2\theta/3+2\pi/3)], & \pi/2\leqslant\theta\leqslant 2\pi\text{'} \end{cases} \tag{15}$$

and (c) mouth pressure analog (Fig. 2, bottom), using the expression

$$y(\theta)=2.18\sum_{n=0}^{10} \mathrm{Real}(C_n)\cos(n\theta)+\mathrm{Imag}(C_n)\sin(n\theta), \tag{16}$$

where C_n are the coefficients in Table I. These values were obtained by first synthesizing a mouth pressure analog signal using the technique described by Titze and Liang (1993), and then extracting its first 11 Fourier coefficients. Coefficient 2.18 sets its peak amplitude to 1.

B. Processing and results

All the algorithms were implemented in Matlab, and run on a personal computer. Trains of cycles of the three signals were synthesized using a sampling frequency of 10 kHz and center frequency $f_0=150$ Hz. Wide ranges of values for K_a and K_f were considered, from 0 (no irregularity) to $K_a=0.1$ and $K_f=0.6$ (large irregularity, as assessed visually). Twenty individual cycles were then identified as the portions with $2\pi(n-1)\leqslant\theta<2\pi n$, $n=1,...,20$, and arranged as a set of 20 wavelets (this number of wavelets was used to insure statistical validity of the results yet to keep the processing task manageable). Each set of wavelets was then normalized in time applying zero-padding, linear normalization (using cubic spline interpolation; Press *et al.*, 1992), and nonlinear normalization using the above FDA algorithm. For comparison, we also applied the Dynamic Programming algorithm given by Qi (1992). Let us briefly recall that this algorithm computes a warping function by minimizing the total square error (cost) between the aligned wavelet and a template, with higher and lower limits imposed to the warping as constraints. The template is chosen as the wavelet with the minimum total cost compared to all the other wavelets. After aligning all wavelets, the mean total cost is the noise energy [denominator in Eq. (8)] divided by a factor of N.

For FDA normalization, we we set function $\alpha(t)$ in Eq. (5), to a constant value of 1000 (i.e., giving the same weight to alignment along all the wavelet length). The constant value was selected so as to have cost function values in the range 1–100, to facilitate the application of the optimization Matlab routine (BFGS quasi-Newton algorithm; Press *et al.*, 1992) available in its standard toolboxes. The roughness penalty parameter was set at $\lambda=1$. This value was selected by visual inspection of the results, so as to allow alignment of the wavelets without significant waveform distortion.

After the normalizations, we computed the HNR and, in the case of nonlinear normalization with FDA, the index of nonlinear warping (INW). Also, we computed the signal-to-noise ratio of the train of cycles

$$\mathrm{SNR}=\frac{\int_0^T y^2[\theta(t)]dt}{\int_0^T \{x(t)-y[\theta(t)]\}^2\,dt}, \tag{17}$$

where $x(t)$ and $y(t)$ are the signals before and after the addition of noise (amplitude variability), respectively [see Eq. (13)], and T is the train's length. Since the SNR measures the ratio of the energy of the signal without noise to the noise energy, it may be regarded as the ''true'' HNR value. Hence, we consider computed HNRs that are closer to the SNR to be more accurate. Let us remark here that the objective of nonlinear normalization is to obtain an HNR value that is close to the SNR [as defined in Eq. (17)] by aligning the wavelets while keeping their general shape.

TABLE I. Fourier coefficients for Eq. (16).

n	C_n
0	$-0.000\,10+i0$
1	$0.068\,83+i0.011\,02$
2	$0.049\,44-i0.078\,75$
3	$-0.031\,22-i0.095\,82$
4	$-0.170\,36-i0.049\,38$
5	$0.029\,29+i0.073\,95$
6	$0.022\,15+i0.020\,59$
7	$0.034\,19-i0.000\,90$
8	$-0.001\,43-i0.005\,95$
9	$-0.000\,56-i0.001\,16$
10	$-0.000\,24-i0.000\,29$

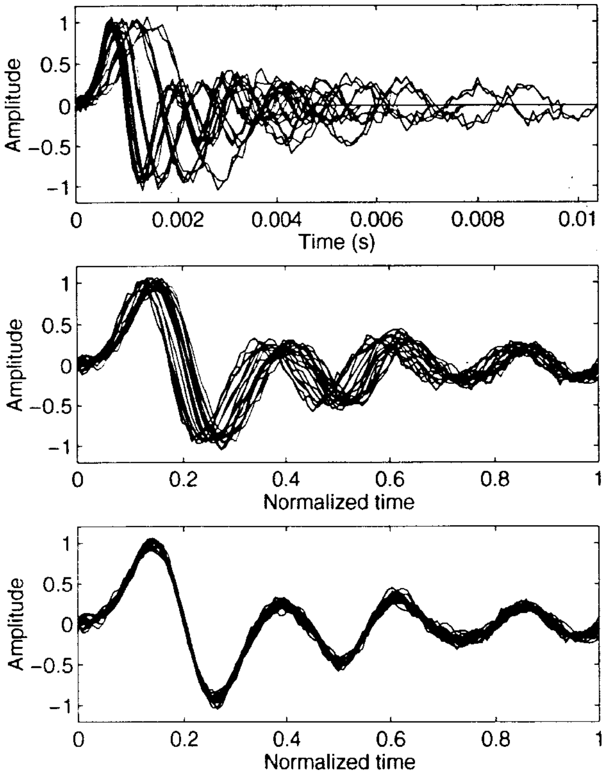

FIG. 3. Mouth pressure wavelets for $K_a = 0.05$ and $K_f = 0.4$, after zero-padding normalization (top), linear normalization (middle), and nonlinear normalization (bottom).

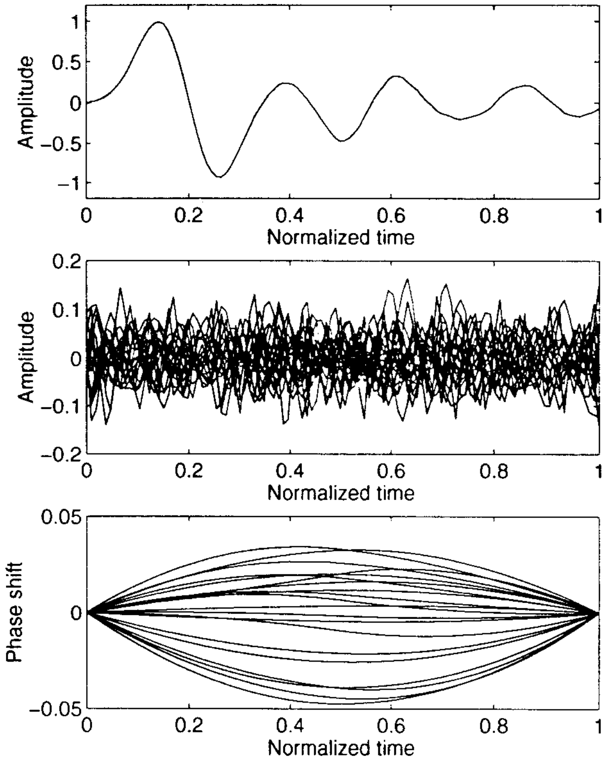

FIG. 4. Average component (top), noise components (middle), and warping functions (bottom) of wavelets in Fig. 3, with nonlinear normalization.

Figure 3 shows an example of the mouth pressure wavelets [Eq. (16)] for $K_a = 0.05$ and $K_f = 0.4$, after zero-padding normalization (top), linear normalization (middle), and nonlinear normalization with FDA (bottom). We can see that nonlinear normalization aligns the wavelets by removing all phase variability. Figure 4 shows the extracted average $\overline{x^*}(t)$ after nonlinear normalization (top), the noise components of the normalized signals $x_i(t) - \overline{x^*}(t)$ (middle), and the phase shift functions $\Delta h_i(t)$ (bottom). The average matches the common pattern of the wavelets, and the noise is uniformly distributed along the wavelets. The HNR is −2.76 dB with zero padding, 6.69 dB with linear normalization, 15.0 dB with nonlinear normalization using Dynamic Programming, and 18.54 dB with nonlinear normalization using FDA. The SNR is 17.82 dB. Clearly, the HNR with nonlinear normalization using FDA produces the best approximation to the SNR of the four methods.

It is instructive to compare the resultant waveforms produced by the two nonlinear normalization methods. Figure 5 shows one of the wavelets after nonlinear normalization by both methods, and the phase shift functions $\Delta h_i(t)$. We can see that the normalized wavelet with FDA maintains the same original shape with a smooth phase shift. On the other hand, the normalized wavelet with Dynamic Programming has noticeable distortions (e.g., compare the shapes of peaks and valleys). These distortions are the consequence of an irregular (nonsmooth) phase shift function.

Figure 6 shows results when the frequency variability of the signal is fixed to $K_f = 0.4$, and the amplitude variability is

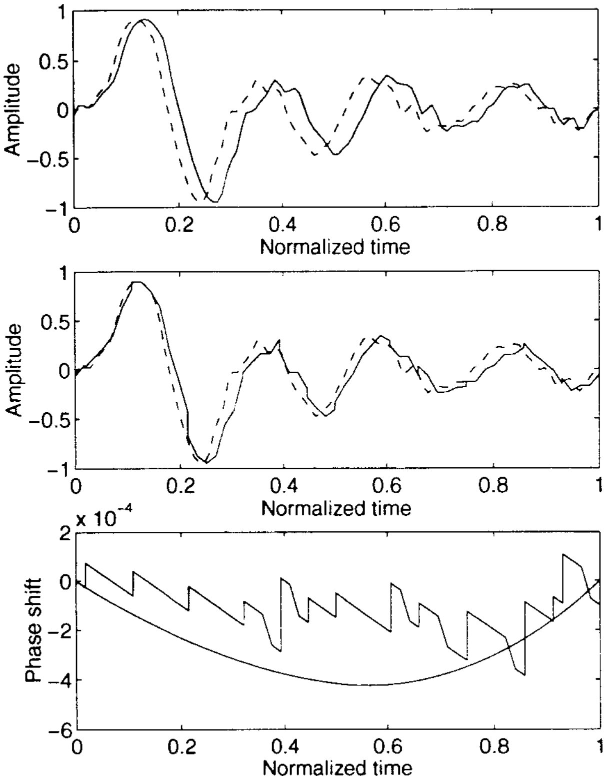

FIG. 5. Comparison of nonlinear normalization results using FDA and Dynamic Programming. Top: linearly normalized wavelet (dashed line) and normalized wavelet using FDA (solid line). Middle: linearly normalized wavelet (dashed line) and normalized wavelet using Dynamic Programming (solid line). Bottom: phase shift functions produced by FDA (smooth curve) and Dynamic Programming (discontinuous curve).

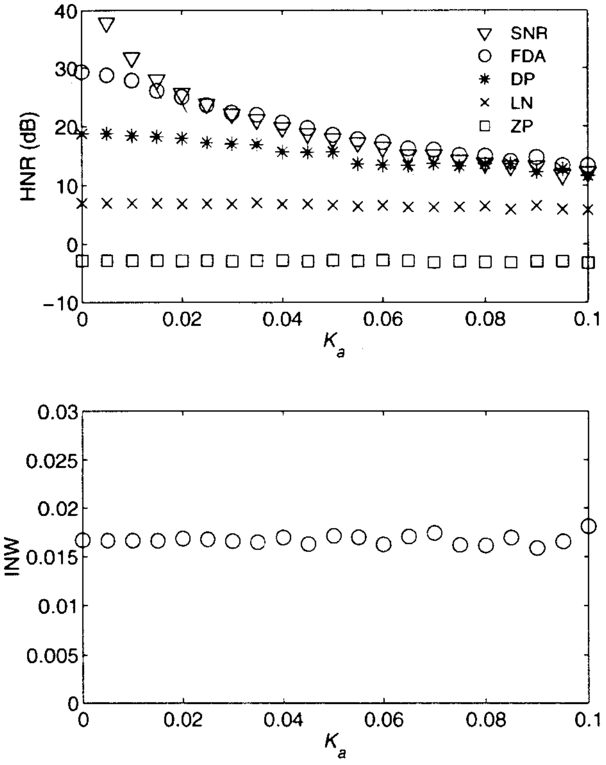

FIG. 6. HNR (top) and INW (bottom) versus amplitude variability K_a, for $K_f=0.4$. Circles: nonlinear normalization using FDA. Stars: nonlinear normalization using Dynamic Programming. Crosses: linear normalization. Squares: zero padding. Triangles: SNR.

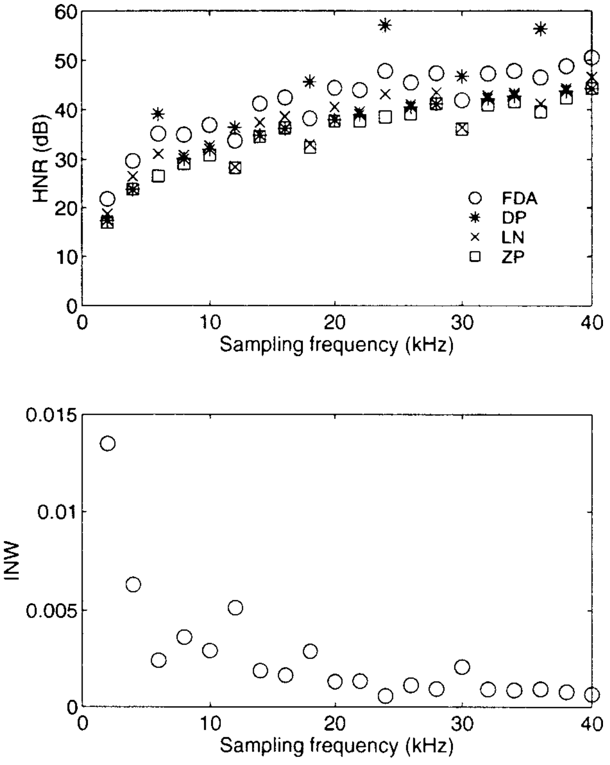

FIG. 7. HNR (top) and INW (bottom) versus sampling frequency, for $K_a=K_f=0$. Circles: nonlinear normalization using FDA. Stars: nonlinear normalization using Dynamic Programming. Crosses: linear normalization. Squares: zero padding. The ideal values of the HNR and INW should be HNR=SNR=∞ and INW=0.

in the range $K_a=0$ to 0.1. The nonlinearly normalized HNR using FDA measures the SNR with good accuracy. The HNR using Dynamic Programming produces results similar to FDA at large values of K_a, but the results worsen as K_a decreases. Both the zero-padded HNR and the linearly normalized HNR are almost constant, since they measure mostly the frequency and phase variability of the wavelets, which is also constant. The lower plot shows the index of nonlinear warping versus the amplitude variability. It is approximately constant in the whole range, reflecting the constant frequency variability.

Figure 6 (top) also shows that, as the amplitude variability tends to 0, the SNR tends to infinity. However, the HNR with both methods of nonlinear normalization reaches a maximum finite value at $K_a=0$. This is a consequence of errors introduced by discretization at the sampling frequency, and numerical errors produced by the algorithms. To assess the degree of precision of the algorithms and test the effect of the sampling frequency, we set both K_a and K_f to zero, and varied the sampling frequency. Figure 7 shows the computed HNR for the four methods and the index of nonlinear warping. In general, as the sampling frequency increases, the HNR increases and INW decreases (ideally, they should be infinite and zero, respectively). The dips and peaks of the HNR and peaks of INW correspond to integer relations between the sampling frequency and central frequency of the signal. In general, the nonlinear normalization with FDA produces more accurate values of the HNR than the other methods. The only exceptions occur at integer relations between sampling frequency and signal central frequency, where nonlinear normalization using Dynamic Programming produces higher values.

Figure 8 shows results when the amplitude variability of the signal is fixed to $K_a=0.05$, and the frequency variability is in the range $K_f=0-0.6$. Both nonlinear normalization methods yield an HNR that is a good approximation to the almost constant SNR, and in general, nonlinear normalization with FDA is more accurate than Dynamic Programming. Both the zero-padded and linearly normalized HNR decrease as the frequency and phase variability increase. We also observe that the index of nonlinear warping increases with the frequency variability, as required.

The results with sine waves and EGG analogs are similar to the ones shown for the mouth pressure analog. According to these results, the nonlinearly normalized HNR using FDA predicts the SNR of the signals with good accuracy. Further, results using FDA are better than results using Dynamic Programming, in the sense that the HNR is usually more accurate and the resultant normalized wavelets and phase shift functions are smoother with FDA, whereas Dynamic Programming introduces significant shape distortions in the wavelets. The results also show that nonlinear normalization is less sensitive to frequency variability than linear normalization or zero padding, thus reducing the effect of jitter on the HNR.

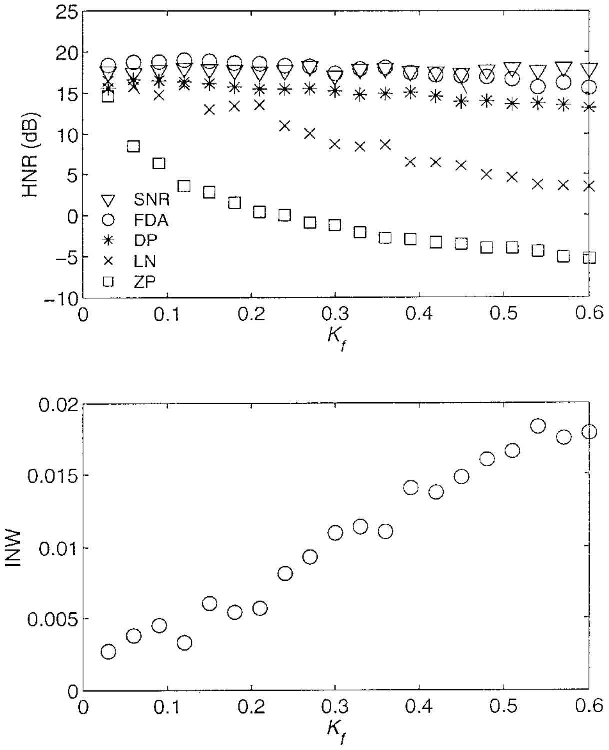

FIG. 8. HNR (top) and INW (bottom) versus frequency variability K_f, for $K_a = 0.05$. Circles: nonlinear normalization using FDA. Stars: nonlinear normalization using Dynamic Programming. Crosses: linear normalization. Squares: zero padding. Triangles: SNR.

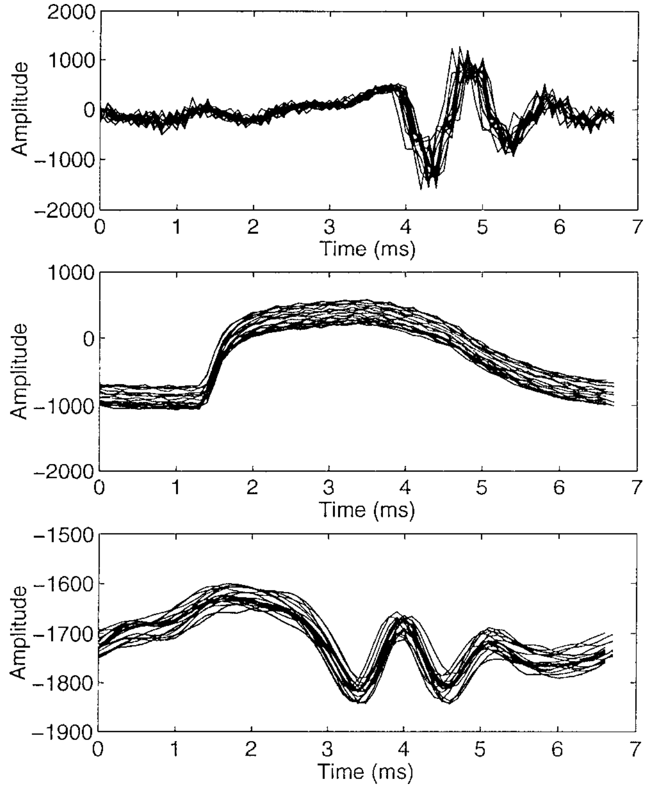

FIG. 9. Recorded acoustic (top), EGG (middle), and airflow (bottom) wavelets for normal speaker A.

IV. EXAMPLES WITH RECORDED VOICE SIGNALS FROM NORMAL SPEAKERS

A. Signals

We next tested the FDA normalization with recorded voice signals. We collected simultaneous acoustics, EGG, and oral airflow from two normal adult subjects producing a sustained /a/. One of the subjects (A) was female, age 33, and the other (B) was male, age 28. The airflow was recorded using an undivided (oral–nasal) Rothenberg mask, and a Glottal Enterprises MSIF-2 filter. The acoustics was recorded with a Seimheiser MKH 816T directional microphone placed outside the mask. The EGG was recorded with a Synchrovoice Research Electroglottograph and Glottal Enterprises Linear Phase Filter and Digital Delay LPHP-2, with settings at 3-kHz frequency limit, no delay, no coupling, and high-pass filter at 5 Hz. All signals were low-pass filtered at 4.8 kHz and digitized at 10 kHz with 12-bit precision. Using the same sampling rate for all signals facilitated the application of the normalization algorithms.

For each subject, the signals were inspected using a signal visualization program, and a stable segment (one which showed the smallest level of amplitude and pattern variability through all cycles, as assessed by visual inspection) was identified, from which 20 consecutive wavelets were extracted from all the three signals. That is, we extracted 3 simultaneous sets (acoustics, EGG, and airflow) of 20 wavelets each. The wavelet boundaries were determined on the EGG using the method of zero crossings with low-pass filtering (Titze and Liang, 1993). Nonlinear normalization was next applied using weights $\alpha = 0.001$, 0.01, and 0.01, for the acoustic, EGG, and airflow wavelets, respectively, and $\lambda = 0.1$. These values were selected following the same criteria as in the case of the synthetic signals. The lower value of α for the acoustic wavelets is a consequence of their larger amplitude values. Selecting a proper value of α might be facilitated by normalizing all wavelets in amplitude prior to the nonlinear normalization, e.g., by dividing the wavelet amplitudes by their peak amplitudes (Wang and Gasser, 1997).

Figure 9 shows the three sets of wavelets for subject A. The three sets have some phase variability, apparently larger in the acoustic and airflow wavelets. We can also note a large dc component in the EGG wavelets, probably produced by vertical movements of the larynx during the recording.

Prior to the FDA normalization, the wavelets were aligned vertically by removing their mean. An alternative for performing a vertical alignment may be to use the first or second derivative of the wavelets (Ramsay and Silverman, 1997). After the derivatives have been normalized in time, then the computed warping functions may be used to normalize the original wavelets. We adopted the first alternative as being computationally simpler, and because the alignment is done directly on the signals whose irregularity is analyzed. The effects of different methods of alignment is a topic that requires further study.

B. Results

Figure 10 shows all normalized wavelets for subject A. The computed indices for all wavelets are listed in Table II. In all cases (except the airflow HNR for subject A), the HNR

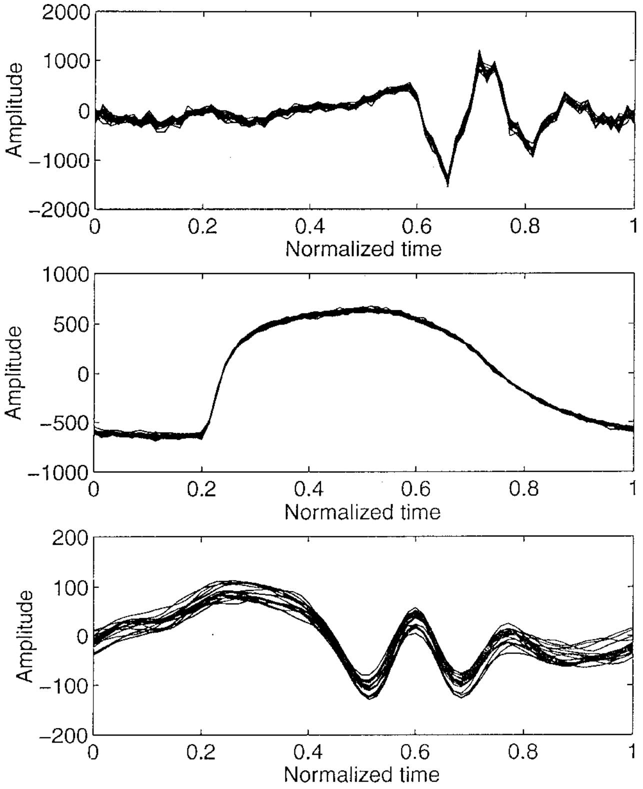

FIG. 10. Normalized acoustic (top), EGG (middle), and airflow (bottom) wavelets for subject A.

values are the highest for nonlinear normalization using FDA, since all phase and frequency variability have been removed by the normalization. We note that the acoustics and oral airflow have similar values of the HNR after nonlinear normalization with FDA, even when the values of HNR using zero padding are a bit different, as in the case of subject B. They also require similar amounts of time warping. A possible interpretation for these results might be that, since the airflow and acoustics are signals produced at the same level (oral output), then one might expect similar indices of amplitude and phase irregularity in both of them. The difference between the acoustic and airflow indices prior to nonlinear normalization would then be an artifact produced by phase shifts between wavelets.

We note also that, in both subjects, the EGG signals have the highest HNR, and require the least warping. The lower values of INW on the EGG suggest that a large phase variability is introduced into the voice above the level of the larynx. This phase variability is produced as a consequence of the vocal tract filtering of the voice source, since different harmonies of the glottal signal are filtered at different gain and timeshift, according to their frequency.

The HNR values with Dynamic Programming are in general a bit lower than values with FDA, except in the case of the airflow for subject A. In the case of EGG for subject A, the HNR results using Dynamic Programming are lower than the value computed with linear normalization. The resultant waveforms show a similar degree of distortion to the example with synthetic signals shown in Fig. 5, which also leads us to question the validity of the HNR measures for the Dynamic Programming normalization. Recall that the objective of nonlinear normalization with FDA is to obtain an HNR value that is close to the SNR [as defined in Eq. (17)] by aligning the wavelets while keeping their general shape. FDA preserves the wavelets' shapes by introducing a roughness penalty constraint in the algorithm [second integral in Eq. (5)]. Dynamic Programming, on the other hand, minimizes the square error measure [first integral in Eq. (5)] only. Since this measure appears in the denominator of the HNR's definition [Eq. (9)], we may then state that Dynamic Programming uses the maximization of the HNR as the criterion for alignment (see also Qi, 1992). However, a higher value of the HNR does not necessarily mean that the value is more accurate. Higher values of the HNR may also be produced by FDA nonlinear normalization, if lower values of the roughness penalty coefficient λ are adopted, at the cost of distorting the wavelets' shapes.

C. Simultaneous normalization

In the previous results, each set of wavelets (acoustics, EGG, and airflow) was normalized separately. As a result, the normalized wavelets are no longer synchronized in normalized time. To keep their synchrony in normalized time, one must perform a simultaneous normalization.

For this, we used vector-valued wavelets, as explained in Sec. II B. Each wavelet was considered three-dimensional, where the three dimensions correspond to the acoustic, EGG, and airflow components. The simultaneous normalization

TABLE II. Computed HNR values for recorded voice signals in normal speakers. ZP: zero-padded HNR; LN: linearly normalized HNR; DP: nonlinearly normalized HNR using Dynamic Programming; FDA: nonlinearly normalized HNR using FDA; INW: index of nonlinear warping.

Signal	ZP (dB)	LN (dB)	DP (dB)	FDA (dB)	INW
		Subject A, female, age 33, $F_0=148.5$ Hz,[a] Jitter=1.1 Hz[b]			
Acoustics	6.7	7.5	10.8	15.9	0.0063
EGG	21.3	27.2	26.8	30.9	0.0030
Airflow	10.6	10.8	13.4	12.4	0.0062
		Subject B, male, age 28, $F_0=109.8$ Hz,[a] Jitter=0.9 Hz[b]			
Acoustics	−1.7	−0.6	9.2	13.7	0.015
EGG	19.6	24.3	27.1	28.6	0.0054
Airflow	8.6	9.1	15.7	15.9	0.017

[a] Mean of $1/T_i$, $i=1,...,20$, where T_i are the wavelet lengths.
[b] Standard deviation of $1/T_i$.

TABLE III. Characteristics of signals from Kay Elemetrics Voice Disorders Database (Kay Elemetrics, 1994).

Speaker	Filename	Age	Sex	Diagnosis	Characteristics	Jitter[a]	Shimmer[a]	NHR[a]	DSH[a,b]
SLM	SLM27AN.NSP	20	M	hyperfunction; anterior–posterior squeezing and ventricular squeezing: head trauma; unilateral paralysis; 7 days post-intubation	moderate shimmer	2.525	14.26	0.233	0
TPS	TPS1GAN.NSP	39	M	unilateral paralysis	moderate shimmer	2.241	14.941	0.25	0
JJD	JJD29AN.NSP	23	M	gastric reflux; bilateral pachydermia and edema; unilateral suleus vocalis	moderate jitter, shimmer, subhar monic components	5.528	12.253	0.411	18.75
VMS	VMS04AN.NSP	27	F	hyperfunction; ventricular compression; bilateral laryngeal web; post-laser removal of subglottic web; scarring	moderate jitter, shimmer, subhar monic components	6.354	15.04	0.495	26.316

[a] Values from the Kay Multidimensional Voice Program.

[b] Degree of subharmonics: estimated relative evaluation of subharmonic to f_0 components in the voice sample.

was performed using the same weights for each component as in the separate normalization.

With simultaneous normalization, the acoustic, EGG, and airflow HNRs for subject A become 11.8 dB, 30.3 dB, and 11.6 dB, respectively, and INW is 0.0039. For subject B, the HNRs are 6.9 dB, 25.9 dB, 11.1 dB, and INW is 0.0087. Comparison with Table II shows that the HNR values are now lower than those for independent normalization, especially for the acoustic wavelets. The lower values of HNR result because the EGG requires a much smaller nonlinear warping than the other two sets (if the EGG required the same amount of warping, both in magnitude and time distribution, as the other two signals, then the results would be similar to the separate normalization). The resultant warping is then a compromise for the three sets. The differences in the results obtained using simultaneous and separate normalization confirm the conclusion obtained from separate normalization, i.e., that a large phase variability is introduced above the larynx.

When only the acoustic and airflow wavelets are simultaneously normalized, their HNRs for subject A are 15.8 dB and 12.1 dB, respectively, with an INW of 0.0061. For subject B we obtain 11.4 dB, 15.7 dB, with an INW of 0.0016. All of these HNR values are now close to those in Table II. This fact shows that the irregularity of the acoustic and airflow wavelets is not only similar (because in Table II the HNR and INW values of the acoustic and airflow signals are similar), but it is also equally distributed in time for both sets (because separate and simultaneous normalization produce similar results).

V. EXAMPLES WITH PATHOLOGICAL VOICE SIGNALS

A. Signals

Finally, we tested the algorithms with voice signals from the Voice Disorders Database of the Voice and Speech Laboratory of the Massachusetts Eye and Ear Infirmary (Kay Elemetrics, 1994). The recording procedure used in the database was as follows: each subject was asked to produced a sustained /a/ at comfortable fundamental frequency and intensity levels. The signals were recorded in a soundproof booth, using a condenser microphone and a DAT-recorder set to a sampling rate of 44.1 kHz. From the DAT-tape the recordings were converted into an analog signal and digitized into a computer at a sampling rate of 25 kHz, with 12-kHz anti-aliasing filtering, and 16-bit resolution.

We selected two signals with high amplitude irregularity but relatively low frequency irregularity (subjects SLM and TPS) and two with high values of jitter and high subharmonic content (subjects JJD and VMS). Table III provides information on the speakers and the main characteristics of their voice signals. The signals were inspected as before using a signal visualization program, and a stable segment was identified, from which 20 consecutive wavelets were extracted. The wavelet boundaries were determined by identifying an easily recognizable event in the signals. We selected the negative zero crossing immediately before the main negative peak in subjects SLM, TPS, and VMS, and the positive zero crossing immediately before the main positive peak in JJD. Other techniques for wavelet extraction were also applied, as discussed in the next section. Nonlinear normalization was applied using a weight $\alpha=10^{-5}$ and $\lambda=0.1$.

B. Results

The computed indices for all signals are listed in Table IV. In general, the FDA nonlinear normalization algorithm achieved a good alignment of all signals, in spite of their high degree of irregularity (see results for subject SLM in Fig. 11) We note that Dynamic Programming produces higher values of HNR than FDA in all cases. One could interpret this result as an indication of better alignment of the wavelets, but a closer look at the resultant waveforms shows a large distortion (see Fig. 12). In this figure, the distortion can be seen at the first positive and negative peaks, and in the flat portion near the end of the wavelet. Looking at the phase shift functions, we see that Dynamic Programming tends to align the wavelet in its finer details, without constraints for smoothness, whereas FDA tends to align its general shape with a smooth phase shift function [the degree of smoothness of the phase shift function can be manipulated by varying

TABLE IV. Computed HNR values for voice signals from the Voice Disorders Database of the Voice and Speech Laboratory, Massachusetts Eye and Ear Infirmary (Kay Elemetrics, 1994). Wavelet boundaries at zero crossing before main positive or negative peak. ZP: zero-padded HNR; LN: linearly normalized HNR; DP: nonlinearly normalized HNR using Dynamic Programming; FDA: nonlinearly normalized HNR using FDA; INW: index of nonlinear warping.

Subject	F_0 (Hz)[a]	Jitter (Hz)[b]	ZP (dB)	LN (dB)	DP (dB)	FDA (dB)	INW
SLM	80.4	2.3	9.1	8.6	13.5	10.6	0.0072
TPS	112.5	1.3	14.8	14.0	19.7	16.5	0.0041
JJD	135.4	5.2	2.7	2.9	7.6	4.7	0.0084
VMS	294.0	8.4	2.1	2.9	7.3	4.4	0.0135

[a] Mean of $1/T_i$, $i=1,\ldots,20$, where T_i are the wavelet lengths.
[b] Standard deviation of $1/T_i$.

coefficient λ in Eq. (5)]. Thus we believe that the higher HNR values of Dynamic Programming are an artifact of the distortion and they do not reflect the actual wavelet irregularity.

Signals of subjects JJD and VMS presented some difficulty due to their content of subharmonies. The frequency spectrum of JJD (computed on the raw signal prior to wavelet extraction) revealed a main peak at 136.4 Hz, with a low frequency component at 68.2 Hz (period-2 phonation; Titze, 1994b). VMS had a main peak at 293.7 Hz, with a lowest frequency component at 59.5 Hz (period-5 phonation; Titze, 1994b). The FDA time normalization performed well when the wavelets were extracted at the higher frequency values (i.e., 136.4 Hz and 293.7 Hz for JJD and VMS, respectively). The indices reported in Table IV correspond to this case. Figure 13 shows results for VMS. On the other hand, at the lower frequency values (i.e., 68.2 Hz and 59.5 Hz for JJD and VMS, respectively), the extracted wavelets presented complex waveform patterns with several peaks and valleys, and the nonlinear normalization algorithm failed to extract a good average. This difficulty might be worst in cases of phonation at two or more incommensurate frequencies (e.g., biphonation) since there is not a consistent waveform pattern repeating at regular intervals, and even identifying the wavelet boundaries would not be trivial. For these cases, voice irregularity may be better evaluated using other techniques, such as those developed by Herzel *et al.* (1994) applying

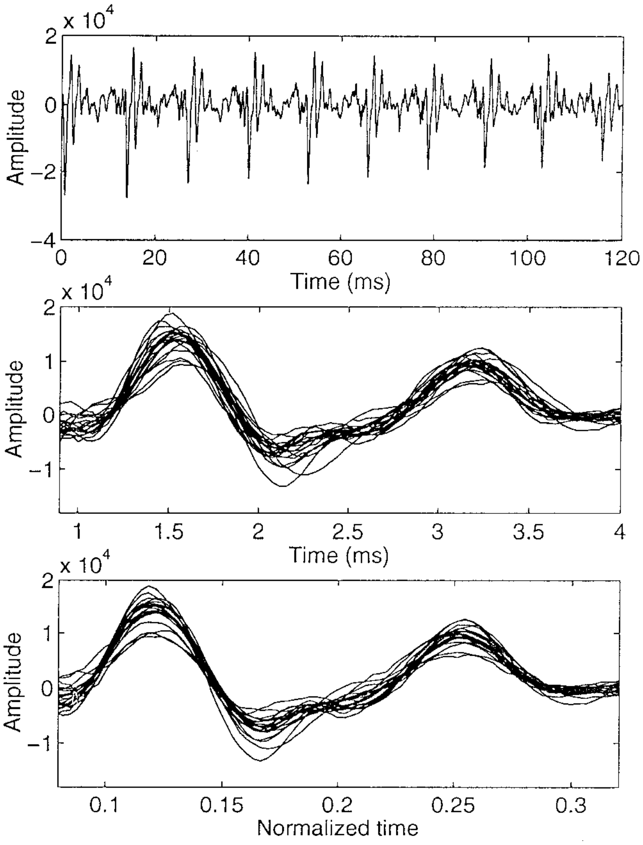

FIG. 11. Original signal (top), extracted unnormalized wavelets (middle), and normalized wavelets using FDA (bottom) for subject SLM. Only a portion of the wavelets is shown, for better visualization of the alignment.

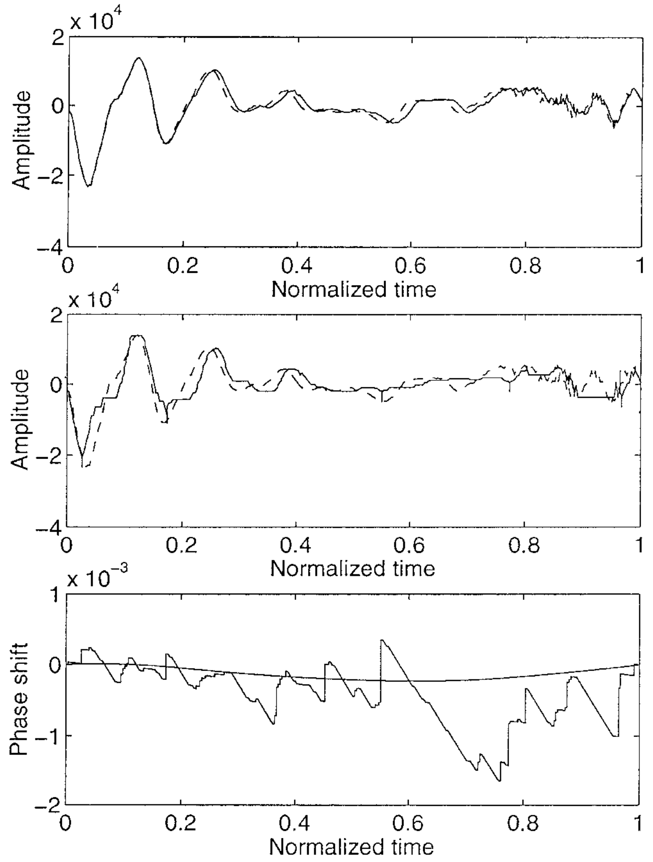

FIG. 12. Comparison of nonlinear normalization results using FDA and Dynamic Programming for subject SLM. Top: linearly normalized wavelet (dashed line) and normalized wavelet using FDA (solid line). Middle: linearly normalized wavelet (dashed line) and normalized wavelet using Dynamic Programming (solid line). Bottom: phase shift functions produced by FDA (smooth curve) and Dynamic Programming (discontinuous curve).

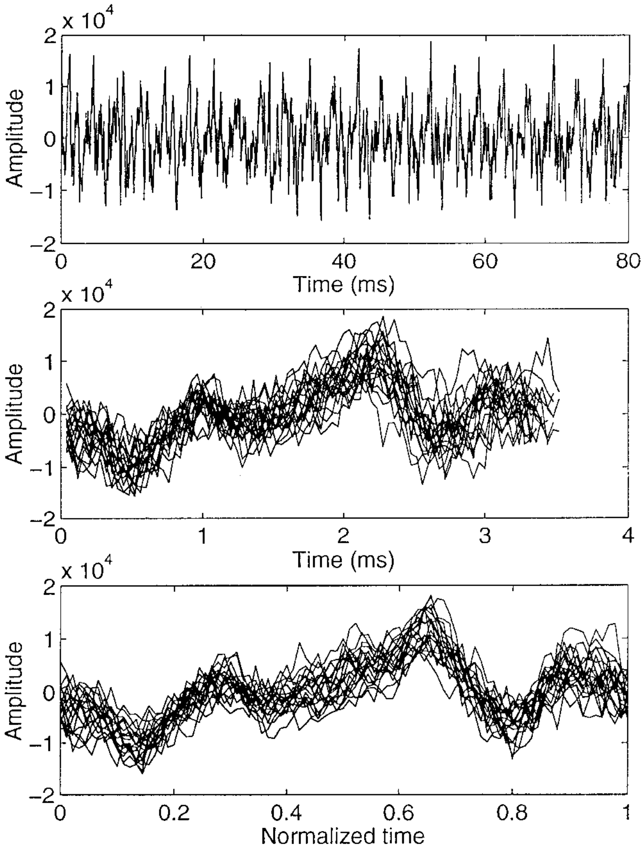

FIG. 13. Original signal (top), extracted unnormalized wavelets (middle), and normalized wavelets (bottom) for subject VMS (wavelet extraction at 293.7 Hz).

nonlinear dynamic theory (bifurcation models, F_0 and amplitude contours, phase portraits, next amplitude and next period maps).

The difficulty encountered with the pathological voices led us to experiment with different techniques to extract the wavelets from the signals. As an example, Table V shows results for negative peak-picking on SLM, TPS, and VMS; and positive peak-picking for JJD. Comparing these values with those in Table IV, there is a mean variation of 2.65 dB, 0.48 dB, 0.35 dB, and 0.15 dB for the zero-padded HNR, linearly normalized HNR, nonlinearly normalized HNR using Dynamic Programming, and nonlinearly normalized HNR using FDA. Wavelet extraction at negative zero crossings of a low-pass filtered signal, and using waveform matching (Titze and Liang, 1993) produced similar results for signals SLM and TPS, and much lower values of HNR for JJD and VMS. In all cases also the variation for both methods of nonlinearly normalized HNR was smaller than for the zero-padded and linearly normalized HNR. These results show that nonlinear normalization is in general less sensitive to the wavelet extraction technique applied in that differences in extraction method yield less HNR variation for nonlinear normalization than for zero-padding or linear normalization. The choice of a wavelet extraction technique seems to be more critical when dealing with highly irregular signals. However, we believe that this is an issue that needs further consideration. For example, it might be possible to combine a waveform matching method of wavelet extraction (which is based on comparing shapes of adjacent cycles) with a wavelet time normalization to improve the selection of the optimal wavelet boundaries. Also, the method used here fixes both ends of wavelets and does not allow any phase shift there. It might be possible to remove the constraints on alignment at the wavelet ends by including additional signal samples before and after the extracted cycles. In this way, errors in detecting the exact cycle boundaries would have less effect on the results.

VI. CONCLUSIONS

We have presented an application of FDA to the time normalization of voice signals and assessment of signal irregularity, which offers certain possible advantages over previous approaches. FDA normalizes the signals while preserving meaningful features of their shapes. Although normalization is done by removing phase differences from the signals, those differences are kept as separate measures in the warping functions. The underlying pattern and irregularity of the signals may then be extracted as separate functions, and the irregularity may be evaluated in terms of irregularity in the waveform (as measured by the HNR), and in phase (as measured by the index of nonlinear warping). We believe that the two indices permit a better assessment of the signal irregularity than a single general index combining phase and waveform irregularity (as the HNR with zero-padding or linear normalization). Two sets of wavelets with different shapes may have the same waveform and phase irregularity, but produce very different values of HNR with zero-padding or linear normalization (as in the examples with the recorded

TABLE V. Computed HNR values for voice signals from the Voice Disorders Database of the Voice and Speech Laboratory, Massachusetts Eye and Ear Infirmary (Kay Elemetrics, 1994). Wavelet boundaries at main positive or negative peak. ZP: zero-padded HNR; LN: linearly normalized HNR; DP: nonlinearly normalized HNR using Dynamic Programming; FDA: nonlinearly normalized HNR using FDA; INW: index of nonlinear warping.

Subject	F_0 (Hz)[a]	Jitter (Hz)[b]	ZP (dB)	LN (dB)	DP (dB)	FDA (dB)	INW
SLM	81.3	2.0	5.3	9.1	13.7	10.9	0.0055
TPS	112.4	1.2	8.9	14.9	19.5	16.4	0.0033
JJD	135.3	3.4	2.0	2.8	7.4	4.7	0.0096
VMS	293.2	9.2	1.9	3.3	8.1	4.2	0.0090

[a] Mean of $1/T_i$, $i=1,...,20$, where T_i are the wavelet lengths.
[b] Standard deviation of $1/T_i$.

airflow and acoustic signals). The similarities in their irregularity would only be revealed after separating them into waveform and phase irregularity.

The possibility of partitioning signal variability into phasing and waveform components also holds promise for research into the nature and significance of variability in speech production. In particular, we plan to extend this method to study of children's speech. Increased variability in children's speech relative to adults has frequently been found (e.g., Chermak and Schneiderman, 1986; Eguchi and Hirsh, 1969; Kent and Forner, 1980; Ohde, 1985; Sharkey and Folkins, 1985; Smith, 1994, 1995; Smith *et al.*, 1983; Tingley and Allen, 1975; Watkin and Fromm, 1984), but the significance and nature of adult–child differences remain matters of debate (see, e.g., Chermak and Schneiderman, 1986; Sharkey and Folkins, 1985; Smith, 1994; Stathopoulos, 1995). More detailed information about how variability is distributed within child and adult data may provide greater insight into the processes by which speech production skill develops.

As shown with synthetic signals, the waveform irregularity measured by nonlinear normalization is less sensitive to jitter, and to errors introduced by sampling frequency discretization than zero-padding or linear normalization techniques. The results from recorded signals show that the irregularity measure is also less sensitive to the wavelet extraction technique used.

We have also shown the limitations of nonlinear normalization using Dynamic Programming. The algorithm tested (Qi, 1992) produced significant distortion in the wavelets, as a consequence of nonsmooth warping functions. Some techniques have been proposed in the literature to reduce wavelet distortion (e.g., Parsons, 1987; Strik and Boves, 1991) by constraining excessive expansion or contraction of the time scale. However, the results (wavelets, warping functions, averages) are in general nonsmooth (i.e., nondifferentiable). Differentiability may be a desirable property, for further processing of the results. One may differentiate the warping functions (which represent phase differences between wavelets) to analyze instantaneous frequency irregularity. For vocal fold oscillation, instantaneous frequency is mainly related to tissue stiffness (Titze, 1994a), so that frequency irregularity (e.g., in the EGG signals) may reveal aspects of the tissue biomechanies and voice motor control. The ability to differentiate a signal one or two times also has great potential applications in work on speech kinematics and aerodynamics. Zero crossings in the first time derivative of a signal may be used as a means of determining the timing of articulatory or aerodynamic events (e.g., Gracco and Löfqvist, 1994; Koenig, in press; Kollia *et al.*, 1995; Löfqvist and Gracco, 1997, 1999). The second time derivative has similarly been used to define articulatery events; for example, Koenig (in press) used the second time derivative in an oral airflow signal to define release and closure for oral stop consonants, and Löfqvist and Gracco (1997) used the second time derivative of a lip opening measure to define the onset of labial closing for a stop consonant. Other potentially interesting FDA techniques also require differentiability of wavelets, such as principal differential analysis, in which a linear differential operator is fitted to the wavelets and irregularity is assessed on a resultant empirical forcing function (see details in Ramsay and Silverman, 1997).

An additional advantage of FDA over Dynamic Programming is that FDA does not require selecting one of the wavelets as a template for the alignment. Further, it allows for considerable flexibility in selecting the alignment criteria, with roughness penalty terms, derivatives, and weighting functions, and one may process simultaneous sets of signals. Simultaneous normalization of signals may be useful for analyzing sets of signals in which one signal is considered to result from variation in other recorded signals. This is the case, for example, with intraoral pressure signals, which vary as a function of both glottal area and supraglottal articulation (e.g., Koenig *et al.*, 1995; Müller and Brown, 1980). Although the warping functions represent a compromise among the various signals, it is possible to vary the influence of various signals so as to achieve an optimal normalization for the particular application.

However, several technical details of the FDA algorithm remain to be further considered, such as how best to select the weighting functions and roughness penalty coefficient, based on the characteristics of the signals. It might also be possible to achieve a better alignment of wavelets by using a weighted combination of the wavelets and their derivatives in the squared error integral in Eq. (5) (Wang and Gasser, 1997), but further study is needed on determining the most appropriate methods for selecting alignment criteria.

In this paper we have used the HNR of the normalized signals as one criterion for selecting among various methods. A noted above, however, obtaining a higher HNR does not always necessarily mean that a method is superior. For example, in the examples with pathological voice signals, time normalization by Dynamic Programming produced higher values of HNR, but at the expense of a large distortion of the wavelets. We do claim that nonlinear normalization using FDA produces a good prediction of the signal SNR [defined as in Eq. (17)], typically higher than the other methods tested, while preserving the shape and smoothness of the resultant wavelets.

In our analyses, we attributed an artificial length of 1 to the normalized wavelet time scale. This is a common technique in FDA, since one is usually interested in analyzing shape characteristics of wavelets. However, it is also possible to interpret the results in absolute time, e.g., by attributing the mean length of the original wavelets to the normalized time length.

The technique is based on the assumption that there is a common pattern to all the wavelets. For regular voices, this assumption is reasonable, and amounts simply to claiming that the laryngeal signal represents the output of a pattern of vocal fold vibration that is essentially periodic, albeit with some minor irregularity due to jitter and shimmer, among other things. The oral signals then represent the periodic laryngeal signal combined with the vocal tract transfer function. Highly disordered voices, on the other hand, present a difficulty for FDA in that the laryngeal signal may not have a consistent pattern repeating at periodic intervals. Cases of signals with subharmonics may still be handled reasonably well by extracting the wavelets at the frequency with the highest energy, but the technique becomes less appropriate as it becomes more difficult to define a single base frequency for the signal.

ACKNOWLEDGMENTS

These results were presented in partial form at the Second International Conference on Voice Physiology and Biomechanics in Berlin, Germany, 12–14 March 1999. We thank Dr. Jean Schoentgen and Dr. Yingyong Qi for helpful discussions; and Dr. Anders Löfqvist, Dr. Helmer Strik, and an anonymous reviewer for comments on a previous version of this paper. This research was supported by CNPq, FEMAT, and CAPES of Brazil, and grants of the US National Institutes of Health (NIH) to Haskins Laboratories.

Chermak, G. D., and Schneiderman, C. R. (**1986**). ''Speech timing variability of children and adults,'' J. Phonetics **13**, 477–480.

Eguchi, S., and Hirsh, I. J. (**1969**). ''Development of speech sounds in children,'' Acta Oto-Laryngol. Suppl. **257**, 1–51.

Gracco, V. L., and Löfqvist, A. (**1994**). ''Speech motor coordination and control: Evidence from lip, jaw, and laryngeal movements,'' J. Neurosci. **14**, 6585–6597.

Herzel, H., Berry, D., Titze, I. R., and Salch, M. (**1994**). ''Analysis of vocal disorders with methods from nonlinear dynamics,'' J. Speech Hear. Res. **37**, 1008–1019.

Kay Elemetrics (**1994**). *Voice Disorders Database, Voice and Speech Laboratory, Massachusetts Eye and Ear Infirmary* (Kay Elemetrics, Lincoln Park, NJ).

Kent, R. D., and Forner, L. L. (**1980**). ''Speech segment durations in sentence recitations by children and adults,'' J. Phonetics **8**, 157–168.

Koenig, L. L. (in press). ''Laryngeal factors in voiceless consonant production in men, women, and 5-year-olds,'' J. Speech Lang. Hear. Res.

Koenig, L., Löfqvist, A., Gracco, V., and McGowan, R. (**1995**). ''Articulatory activity and aerodynamic variation during voiceless consonant production,'' J. Acoust. Soc. Am. **97**, S3401.

Kollia, H. B., Gracco, V. L., and Harris, K. S. (**1995**). ''Articulatory organization of mandibular, labial, and velar movements during speech,'' J. Acoust. Soc. Am. **98**, 1313–1324.

Lieberman, P. (**1962**). ''Some acoustic measures of the fundamental periodicity of normal and pathological larynges,'' J. Acoust. Soc. Am. **35**, 344–353.

Löfqvist, A., and Gracco, V. L. (**1997**). ''Lip and jaw kinematics in bilabial stop consonant production,'' J. Speech Lang. Hear. Res. **40**, 877–893.

Löfqvist, A., and Gracco, V. L. (**1999**). ''Interarticulator programming in VCV sequences: Lip and tongue movements,'' J. Acoust. Soc. Am. **105**, 1864–1876.

Lucero, J. C., Munhall, K. G., Gracco, V. L., and Ramsay, J. O. (**1997**). ''On the registration of time and the patterning of speech movements,'' J. Speech Lang. Hear. Res. **40**, 1111–1117.

Müller, E. M., and Brown, W. S. (**1980**). ''Variations in the supraglottal air pressure waveform and their articulatory interpretation,'' in *Speech and Language: Advances in Basic Research and Practice*, edited by N. Lass (Academic, New York), pp. 317–389.

Ohde, R. N. (**1985**). ''Fundamental frequency correlates of stop consonant voicing and vowel quality in the speech of preadolescent children,'' J. Acoust. Soc. Am. **78**, 1554–1561.

Parsons, T. W. (**1987**). *Voice and Speech Processing* (McGraw-Hill, New York).

Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (**1992**). *Numerical Recipes in C—The Art of Scientific Computing*, 2nd ed. (Cambridge University Press, Cambridge).

Qi, Y. (**1992**). ''Time normalization in voice analysis,'' J. Acoust. Soc. Am. **92**, 2569–2576.

Qi, Y., and Hillman, R. E. (**1997**). ''Temporal and spectral estimations of harmonics-to-noise ratio in human voice signals,'' J. Acoust. Soc. Am. **102**, 537–543.

Qi, Y., Weinberg, B., Bi, N., and Hess, W. J. (**1995**). ''Minimizing the effect of period determination on the computation of amplitude perturbation in voice,'' J. Acoust. Soc. Am. **97**, 2525–2532.

Ramsay, J. O. (**1998**). ''Estimating smooth monotone functions,'' J. Royal Stat. Soc., Ser. B **60**, 365–375.

Ramsay, J. O., and Li, X. (**1998**). ''Curve registration,'' J. Royal Stat. Soc. B **60**, 351–363.

Ramsay, J. O., Munhall, K. G., Gracco, V. L., and Ostry, D. J. (**1996**). ''Functional data analyses of lip motion,'' J. Acoust. Soc. Am. **99**, 3718–3727.

Ramsay, J. O., and Silverman, B. W. (**1997**). *Functional Data Analysis* (Springer-Verlag, New York).

Sharkey, S. G., and Folkins, J. W. (**1985**). ''Variability of lip and jaw movements in children and adults: Implications for the development of speech motor control,'' J. Speech Hear. Res. **28**, 8–15.

Smith, B. L. (**1994**). ''Effects of experimental manipulations and intrinsic contrasts on relationships between duration and temporal variability in children's and adults' speech,'' J. Phonetics **22**, 155–175.

Smith, B. L. (**1995**). ''Variability of lip and jaw movements in the speech of children and adults,'' Phonetica **52**, 307–316.

Smith, B. L., Sugerman, M. D., and Long, S. H. (**1983**). ''Experimental manipulation of speaking rate for studying temporal variability in children's speech,'' J. Acoust. Soc. Am. **74**, 744–749.

Stathopoulos, E. T. (**1995**). ''Variability revisited: An acoustic, aerodynamic, and respiratory kinematic comparison of children and adults during speech,'' J. Phonetics **23**, 67–80.

Strik, H., and Boves, L. (**1991**). ''A dynamic programming algorithm for time-aligning and averaging physiological signals related to speech,'' J. Phonetics **19**, 367–378.

Tingley, B. M., and Allen, G. D. (**1975**). ''Development of speech control in children,'' Child Dev. **46**, 186–194.

Titze, I. R. (**1994a**). *Principles of Voice Production* (Prentice-Hall, Englewood Cliffs).

Titze, I. R. (**1994b**). *Workshop on Acoustic Voice Analysis—Summary Statement* (National Center for Voice and Speech, Iowa City).

Titze, I. R., and Liang, H. (**1993**). ''Comparison of F_0 extraction methods for high-precision voice perturbation measurements,'' J. Speech Hear. Res. **36**, 1120–1133.

Wang, K., and Gasser, T. (**1997**). ''Alignment of curves by dynamic time warping,'' Annals of Statistics **25**, 1251–1276.

Watkin, K., and Fromm, D. (**1984**). ''Labial coordination in children: Preliminary considerations,'' J. Acoust. Soc. Am. **75**, 629–632.

Yumoto, E., Gould, W. J., and Baer, T. (**1982**). ''Harmonies-to-noise ratio as an index of degree of hoarseness,'' J. Acoust. Soc. Am. **71**, 1544–1550.