Discriminating Simulated Vocal Tremor Source Using Amplitude Modulation Spectra

Kathy M. Carbonell, Rosemary A. Lester, Brad H. Story, and Andrew J. Lotto, *Tucson, Arizona*

Summary: Objectives/Hypothesis. Sources of vocal tremor are difficult to categorize perceptually and acoustically. This article describes a preliminary attempt to discriminate vocal tremor sources through the use of spectral measures of the amplitude envelope. The hypothesis is that different vocal tremor sources are associated with distinct patterns of acoustic amplitude modulations.

Study Design. Statistical categorization methods (discriminant function analysis) were used to discriminate signals from simulated vocal tremor with different sources using only acoustic measures derived from the amplitude envelopes.

Methods. Simulations of vocal tremor were created by modulating parameters of a vocal fold model corresponding to oscillations of respiratory driving pressure (respiratory tremor), degree of vocal fold adduction (adductory tremor), and fundamental frequency of vocal fold vibration (F0 tremor). The acoustic measures were based on spectral analyses of the amplitude envelope computed across the entire signal and within select frequency bands.

Results. The signals could be categorized (with accuracy well above chance) in terms of the simulated tremor source using only measures of the amplitude envelope spectrum even when multiple sources of tremor were included.

Conclusions. These results supply initial support for an amplitude-envelope-based approach to identify the source of vocal tremor and provide further evidence for the rich information about talker characteristics present in the temporal structure of the amplitude envelope.

Key Words: Vocal tremor–Envelope modulation spectra–Acoustic analysis.

INTRODUCTION

Characteristics of vocal tremor

Vocal tremor is a voice disorder that is characterized by an unstable or shaky-sounding voice[1] and measurable modulation of the acoustic output.[2–10] These perceptual and acoustical characteristics are produced by tremor affecting components of the speech mechanism including the respiratory system,[11–13] larynx,[2,3,6,7,9,10,12,14,15] and vocal tract.[2,4,7,11,16–19] Tremor is associated with several different neurologic disorders including essential tremor, Parkinson disease, cerebellar dysfunction, and dystonia.[20] In individuals with essential tremor, the most common tremor disorder, vocal tremor is estimated to occur in approximately 18–30% of cases.[19,21,22]

Previous research on essential vocal tremor has demonstrated that tremor affecting the structures within the speech mechanism produced nearly rhythmic modulation of the fundamental frequency (F0) and the intensity of the voice during sustained vowel production.[2–10] The primary focus of this research was on measuring the modulation rate (ie, the number of cycles of modulation that occur within 1 second) and the modulation extent (ie, the range of modulation) of F0 and intensity. Dromey et al[5] reported that the rate of F0 modulation ranged from 3.2 to 5.3 Hz and, similarly, the rate of intensity modulation ranged from 2.6 to 5.0 Hz during sustained vowels produced at a comfortable pitch and loudness by individuals with essential vocal tremor. The extent of F0 modulation in this study ranged from 2.9 to 15.0%; whereas, the extent of intensity modulation ranged from 18.5 to 55.6%. In a study of respiratory and laryngeal vocal tremor using acoustic analyses and electromyography, Koda and Ludlow[12] found that the mean rate of modulation of the acoustic signal was 4.9 Hz. The rate of the acoustical modulations was consistent with the rate of the measured physiological modulations. That is, the mean rate of modulation of muscle activation of the two primary intrinsic laryngeal muscles involved in F0 control was 4.7 Hz in the thyroarytenoid and 5.1 Hz in the cricothyroid. The mean rate of modulation carried onto the respiratory structures and measured using respiratory inductive plethysmography for the same participants was 4.6 Hz.[12] Measurements of both the rate and the extent of F0 and intensity modulation varied when individuals produced different pitches and loudness levels.[5]

In the majority of studies on essential vocal tremor, either the involvement of each component of the speech mechanism was not identified or multiple components of the speech mechanism were affected by tremor. As a result, it is uncertain whether specific acoustic modulation patterns are associated with tremor affecting the respiratory system, the larynx, or the vocal tract (for a review of possible contributions of each component of the speech mechanism to vocal tremor, see Lester, Barkmeier-Kraemer, and Story[7]).

Different methods have been proposed to improve acoustic analysis of vocal tremor for clinical identification and characterization of the source of vocal tremor including the vocal demodulator[23] and the modulogram.[24] The vocal demodulator measured the extent and rate of F0 modulation and of F0 amplitude modulation, with a range of modulation rate limited to 2.5 to 25 Hz. As an extension of the vocal demodulator, the modulogram analyzed the rate and extent of modulation of F0 and overall amplitude with three distinct rate bands: flutter (10–20 Hz), tremor (2–10 Hz), and wow (0.2–2 Hz). The vocal demodulator and modulogram were both used to analyze vocal tremor, and the modulogram was used to measure change in

Accepted for publication July 31, 2014.

From the Department of Speech, Language, and Hearing Sciences, University of Arizona, Tucson, Arizona.

Address correspondence and reprint requests to Kathy M. Carbonell, Department of Speech, Language, and Hearing Sciences, University of Arizona, P.O. Box 210071, Tucson, AZ 85721-0071. E-mail: kathyc@email.arizona.edu

Journal of Voice, Vol. 29, No. 2, pp. 140-147
0892-1997/$36.00

http://dx.doi.org/10.1016/j.jvoice.2014.07.020

vocal stability pre- and post-botulinum toxin (Botox) injections in individuals with vocal tremor.[14] However, it is difficult to clinically apply the results of these studies because these methods were used to analyze vocal tremor in individuals with a variety of neurologic disorders affecting more than one component of the speech mechanism. Without knowing the specific involvement of the respiratory systrynx, and vocal tract in vocal tremor and their direct effect on the acoustic patterns, it continues to be challenging to identify the source of vocal tremor in individuals with various neurologic etiologies.

To our knowledge, only two studies have systematically investigated the acoustic patterns of vocal tremor and the underlying physiology. Jiang et al[25] isolated the source of tremor to subglottal, glottal, and supraglottal levels using an electric tapping device on the chest, side of the throat, and cheek during phonation to simulate vocal tremor in healthy adults. Acoustic analyses included measures of F0, percent jitter (ie, cycle-to-cycle frequency perturbation), and percent shimmer (ie, cycle-to-cycle amplitude perturbation). In addition, the frequency and amplitude contours of the acoustic signal were extracted, and peak prominence ratios were calculated for the contours by determining the energy of each peak and dividing it by the total energy within the signal. The peak prominence ratios derived from the frequency contour distinguished between the normal condition and the three simulated tremor conditions. The peak prominence ratio derived from the amplitude contour distinguished between all pairs of tremor conditions, except the chest and throat conditions (ie, the subglottal and glottal conditions). These results indicated that extraction of amplitude contours to derive power spectra is a useful technique in discriminating different sources of tremor. However, Lester and Story[26] demonstrated that individuals respond to forced oscillation of the respiratory apparatus with adaptations at the level of the larynx (ie, changes in the mean F0 with different magnitudes of applied pressure oscillation). Therefore, a more controlled study of the characteristics of amplitude envelope modulation for isolated sources of tremor was warranted.

In an attempt to completely isolate the source of tremor to the larynx and to determine the associated acoustic characteristics, Lester et al[7] simulated oscillations affecting only fundamental frequency to simulate vocal tremor affecting fold length (throughout this article, we use "source" to refer to the proximal cause of the acoustic perturbations as opposed to the causal "source" of the tremor, which is neural). Simulations were created using a kinematic model of the vocal folds[27,28] coupled to a wave-reflection model of the trachea and a parametric model of the vocal tract area function (J. Liljencrants, Unpublished doctoral dissertation, 1985; B. H. Story, Unpublished doctoral dissertation, 1995).[29] Acoustic analyses of this simulated vocal tremor revealed comodulation of the F0 and overall intensity of the acoustic signal. Spectral analyses of the amplitude envelope were not investigated in this study nor have they been conducted with simulated vocal tremor affecting other parts of the speech mechanism.

In summary, vocal tremor can be caused by isolated or combined sources of tremor within the speech mechanism. Each source may produce modulation of the F0 and intensity of the resultant waveform. It is unclear whether measures of the rate or extent of acoustic modulation can distinguish the source of vocal tremor. However, spectral analyses of the amplitude envelope may provide discriminating information about source.[25]

Envelope modulation spectrum

It has long been known that rhythmic modulations of the amplitude envelope in speech are important for intelligibility. Filtering or adding noise to the amplitude envelope in the frequency range of 2–9 Hz substantially decreases intelligibility.[30–32] There has been a recent resurgence in scientific interest in the modulations of the amplitude envelope as it has been suggested that neural oscillations may entrain to the rhythmic modulations of the envelope as a means of tuning perceptual processing to the intrinsic rhythms of speech (eg, study by Ghitza,[32,33] Peelle and Davis,[34] Peelle et al,[35] Luo[36]).

In addition to providing information about the linguistic content of the signal, the rhythmic characteristics of the amplitude envelope in speech may carry information about the talker. Measurements derived from the amplitude envelope have been demonstrated to differ based on the language spoken by a talker (eg, Korean vs Mandarin[37]) and can even distinguish individual speakers within a single language.[38,39] Such measures have also been shown to be effective in distinguishing clinically relevant categories of speech perturbations. Liss et al[40] proposed that the power spectra of the amplitude envelope for speech computed for the entire signal and for octave frequency bands, which they called the envelope modulation spectrum (EMS), would contain relevant information about the rhythmic disruptions that are a common symptom of dysarthria. In support of this proposal, they computed a number of metrics (such as the peak frequency in the spectrum and the relative energy in the spectrum above and below 4 Hz) and demonstrated that one could reliably categorize speakers into subtypes of dysarthria (eg, ataxic, hypokinetic, etc.). In a discriminant analysis using just these metrics, they were able to classify 43 speakers across five subtypes of dysarthria with 84% accuracy.

The success of the EMS approach in classifying dysarthric subtypes suggests that similar amplitude envelope metrics may be able to classify vocal tremor sources from the acoustic signal, especially given that the major acoustic consequence of tremors have a strong temporal structure. To determine the utility of this approach for tremor, we follow the empirical logic of Liss et al[40] using functional discriminant analysis to categorize signals solely on the basis of metrics derived from the amplitude envelope spectra.

Whereas Liss et al[40] had previous dysarthric subtype diagnoses for each of their speakers to use as a basis for determining classification accuracy; it is difficult to provide a definitive diagnosis of vocal tremor source in many patients. Therefore, we followed the approach of Lester et al[7] and used simulations of vocal tremor as the basis for our two experiments. The use of simulated vocal tremor provides us the benefits of (i) knowing with certainty the source of the tremor; (ii) the ability to create tremor with isolated sources or compound sources; and (iii) the ability to independently vary parameters of the tremor such as the rate of modulation, extent of modulation, and mean F0.

A set of sustained vowels were simulated by the computational model (described in the following) with different isolated tremor sources (Experiment 1) or with compound sources (Experiment 2). The EMS measures described by Liss et al[40] were calculated for each synthesized signal. These measures were then entered into a stepwise discriminant function analysis (DFA) with the vocal tremor source serving as the grouping variable. If there is information in the amplitude envelope spectra for distinguishing vocal tremor sources, then one would predict that the classification accuracy of the discriminant analysis will be above chance. On the other hand, if different vocal tremor sources do not result in discriminably different amplitude envelope patterns, one would not expect the discriminant analysis to provide classification accuracy above chance. If the former prediction is supported, it would provide motivation for further exploration of the diagnostic possibilities of acoustic analyses of vocal tremor.

EXPERIMENT 1

Method

Tremor database. The simulations of vocal tremor were generated using a kinematic model of the voice source[27,28] coupled to a wave-reflection model of the trachea and vocal tract.[41] This model allowed for control of the pressure supplied to the larynx, the fundamental frequency (corresponding to vocal fold length), and the degree of vocal fold adduction. This model also allowed for control of the configuration of the vocal tract filter for vowel shaping via a parametric model of the vocal tract area function.[29] In the version of this model used for the present study, the acoustic output was influenced by the resonances of the trachea, as in the human system. The result was high-quality acoustic signals for sustained vowels that were suitable for acoustical analyses.

The kinematic model of the voice source was based on a representation of the medial vocal fold surfaces. When these surfaces were driven in a vibratory pattern, as they are during voice production, a time-varying glottal area was generated. These changes in glottal area interacted with the pressures in the trachea and vocal tract and generated a time-synchronous glottal flow. This glottal flow generated a pressure wave that propagated through the vocal tract and ultimately produced the speech signal radiated at the lips, which is analogous to a microphone signal from a human talker.

The position of the vocal fold surfaces and the characteristics of their vibration in this model were defined by a combination of user input and rules derived from anatomical and physiological studies.[28] The parameter settings for the present study are presented in Table 1. These were selected to simulate a normal vocal system with an isolated oscillation representative of vocal tremor. For further details related to the model parameters, see the study by Story.[41]

TABLE 1.
Parameter Settings of the Kinematic Model of the Vocal Folds Used in this Study

Parameter	Setting
Pressure (dyn/cm^2)	8000
Superior adduction (ξ_{02}) (cm)	0.08
Surface bulging (ξ_b) (cm)	0.05
Nodal point ratio (z_n)	0.8
Epilaryngeal area (cm^2)	0.5

For Experiment 1, simulations of vocal tremor were limited to the modulation of single parameters of the speech production system. This allowed for analysis of the acoustic output of three different sources of tremor: vocal fold adduction (adductory tremor), fundamental frequency of vibration (F0 tremor), and respiratory pressure (respiratory tremor). To generate the set of simulations, degree of adduction, F0, and respiratory pressure were each modulated with three extents (2.5, 5, and 10%) and three rates (3, 5, and 7 Hz). These were imposed on three different settings of both the vocal tract shape and fundamental frequency of vibration. For the vocal tract shape, area functions corresponding to the vowels [a], [i], and [u] were used, whereas the F0 was set to 100, 120, and 140 Hz to roughly approximate the range of an adult male voice. With all combinations, 243 vocal tremor samples were generated.

Any given tremor was simulated by modulating the relevant parameter about its baseline setting at a particular rate and extent. For example, an adductory tremor with a modulation extent of 2.5% would cause the distance between the vocal processes to oscillate between 0.078 and 0.082 cm assuming the baseline value of 0.08 cm given in Table 1. This has the effect of imposing of slowly varying modulation on the glottal area signal generated as the vocal folds vibrate. For an F0 tremor, the modulation is imposed directly on the value of F0; if F0 were set at 100 Hz, a 5% extent would cause the F0 to oscillate between 95 and 105 Hz. A varying F0 also has the effect of decreasing and increasing the vocal fold length, which consequently modulates the glottal area. As an example for respiratory pressure, a 10% modulation would vary the driving pressure in a range from 7200 to 8800 dyn/cm^2 assuming the baseline value of 8000 dyn/cm^2 given in Table 1. The comprehensive summary of the controlled parameters of the synthesized vocal tremor database are presented in Table 2. All samples were saved as mono.wav files with a sampling rate of 44.1 kHz.

The database developed for this test is based on a number of simplifying assumptions (single vocal tract shape, steady state vowel production, etc.) to provide a constrained set of stimuli. Whereas these assumptions are not representative of "real-world" variability, the database provides an initial test of the discriminative abilities of EMS for tremor source. If EMS proves capable of discriminating tremor source, additional variability can be parametrically added to the database to determine how well this variability can be accommodated (as in Experiment 2).

Acoustic measures

The present study performed acoustic analyses on synthesized vocal tremor using metrics of the EMS developed by Liss

TABLE 2.
Summary of the Controlled Parameters of the Synthesized Tremor Database, for a Total of 243 Vocal Tremor Samples

Tremor Type	Extent of Modulation (%)	Rate of Modulation (Hz)	Steady-State Vowel	Fundamental Frequency (Hz)
Adductory	2.5; 5; 10	3; 5; 7	[a]; [i]; [u]	100; 120; 140
F0	2.5; 5; 10	3; 5; 7	[a]; [i]; [u]	100; 120; 140
Pressure	2.5; 5; 10	3; 5; 7	[a]; [i]; [u]	100; 120; 140

Notes: Each tremor sample has a total duration of 3 seconds.

et al.[40] Each set of analyses is performed on the full signal and also on seven octave bands (center frequencies of 125, 250, 500, 1000, 2000, 4000, and 8000 Hz). The inclusion of the octave bands provides amplitude modulation information that is specific for low versus high frequencies, and the information from these bands is often more discriminating than that obtained for the full signal (eg, 42, 41). The first step of the analysis is the extraction of the amplitude envelope (half wave rectification followed by 30-Hz low-pass filtering). The power spectrum is then calculated (512-point fast Fourier transform (FFT)) providing a spectrum of the frequencies up to 10 Hz for the envelope. From this spectrum six dependent variables are calculated. The names and descriptions of these variables are presented in Table 3. The result is a set of 48 dependent variables (eight envelopes $\times$ six variables) for each tremor sample. Further details on the motivation of these variables and their calculation can be found in the study by Liss et al.[40]

Discriminant function analysis

The purpose of this study was to determine whether the temporal regularities present in vocal tremor acoustics could reliably distinguish between independent sources of tremor. The dependent variables derived from EMS served as the discriminating variables in this analysis.

Stepwise DFA was computed for the entire set of synthesized vocal tremors in the database to determine which of the 48 dependent variables would best discriminate among the three different sources of tremor. At each stage of the DFA, the variable that minimized Wilks lambda was entered, provided the F statistic for the change was significant ($P < 0.05$). At any point during the analysis, variables were removed from the DFA if they were found to no longer significantly lower Wilks lambda ($P > 0.10$) when a new variable was added. Canonical functions, representing linear combinations of the selected discriminating variables were constructed by the DFA and used to create classification rules for group membership. The accuracy with which these rules classify the members of the group is expressed as a percentage correct. As a test of the robustness of the classification rules, we used cross-validation (also called the "leave-one-out") method. By this method, the DFA constructs the classification rules using all but one of the tremor samples and then the excluded tremor sample is classified based on the functions derived from all other tremor samples. Following the logic of Liss et al,[40] classification accuracy significantly greater than chance provides evidence that there is information about vocal tremor source present in the amplitude envelope spectra of sustained vowels.

Results

This analysis classified each of the synthesized vocal tremor samples (243 exemplars) as belonging to one of the three types of vocal tremor source: adductory, F0, or pressure. Eleven of the 48 EMS metrics emerged as discriminating variables (Table 4 for a list of all significant variables in the final model). The DFA classification accurately classified 89.7% of the tremor samples as belonging to their designated source groups (86% on cross-validation, chance = 33.3%). A second DFA, using only the top five variables, demonstrated a tremor source classification accuracy of 79.4% (77.8% on cross-validation). The fact that these accuracy values are well above chance provides support for the hypothesis that there is information relevant to

TABLE 3.
The Six Acoustic-Dependent Variables Obtained for From the Amplitude Envelope Spectrum for Each of the Octave Bands and the Full Signal (Following Liss et al[40])

Dependent Variable Measure	Description
Peak frequency	The frequency of the peak in the spectrum with the greatest amplitude. The period of this frequency is the duration of the predominant repeating amplitude pattern.
Peak amplitude	The amplitude of the peak described previously (divided by overall amplitude of the energy in the spectrum). This is a measure of how much the rhythm is dominated by a single frequency.
Energy 3–6	Energy in the region of 3–6 Hz (divided by overall amplitude of spectrum).
Below 4	Energy in spectrum from 0 to 4 Hz (divided by overall amplitude of spectrum).
Above 4	Energy in spectrum from 4 to 10 Hz (divided by overall amplitude of spectrum).
Ratio 4	Below 4/above 4.

TABLE 4.
Results of Independent Tremor Analysis, Classifying 243 Exemplars into Three Tremor Types (Adductory, f0, Respiratory); Chance = 33.3%

Analysis	Classification Accuracy	Top Discriminating Variables	
EMS	89.7% (with 11 variables)	Peak amplitude (125 Hz) Peak amplitude (full signal) Ratio 4 (500 Hz) Energy 3–6 (125 Hz) Below 4 Hz (500 Hz) Peak frequency (500 Hz)	Peak frequency (1000 Hz) Peak amplitude (8000 Hz) Peak frequency (4000 Hz) Peak amplitude (2000 Hz) Below 4 (125 Hz)
EMS	79.4% (top 5 variables only)	Peak amplitude (125 Hz) Peak amplitude (full signal) Ratio 4 (500 Hz) Energy 3–6 (125 Hz) Below 4 (500 Hz)	

vocal tremor source in amplitude modulation spectra for isolated tremor sources.

EXPERIMENT 2

Whereas the results of Experiment 1 demonstrated that amplitude envelope spectra measures can discriminate isolated sources of vocal tremor, the situation in real patients is likely to be much more complex. In particular, multiple sources of vocal tremor are likely to co-occur. It is of practical and theoretical relevance whether there is sufficient information in the EMS metrics to detect a particular source of tremor whether it is present in isolation or in combination with other sources. To examine this question, we created a set of simulated tremor signals that contained each of the three isolated sources from Experiment 1 as well as each of the three possible combinations of two of these sources. We then asked whether DFA could accurately group the signals together that contained an adductory tremor either in isolation or in combination (the choice of adductory as the grouping variable was arbitrary). Classification accuracy that is greater than chance would be taken as evidence that the information about tremor source in the amplitude envelope is robust even in the presence of additional tremor sources.

Method

Tremor samples. The stimuli for Experiment 2 were synthesized using the same kinematic model of the vocal folds attached to a model of the vocal tract and trachea described for Experiment 1. A set of single sources of tremor from Experiment 1 were included in the analyses for Experiment 2. Because of the increase in the number of tremor types, the number of other parameters that were manipulated was decreased. Sources of tremor varied only in the extent of modulation (2.5%, 5%, and 10%). Rate of modulation (5 Hz), steady-state vowel ([a]), duration (3 seconds), and fundamental frequency (100 Hz) were kept constant. In addition to the single source tremors, the three possible two-source combination tremors were simulated as well with the same parameters. The complete set of stimuli is described in Table 5.

It is important to note that a real-world presence of vocal tremor from a combination of sources would likely result in an interaction based on physiological constraints.[26] For the purpose of this study, however, a controlled, noninteracting combination tremor allows us to test the sensitivity of EMS measures in identifying salient temporal regularities present in the acoustic signal, particular to individual sources of tremor.

RESULTS

The goal of this analysis was to test the sensitivity of EMS measures in classifying the presence of an individual source of tremor (ie, adductory), even when presented in combination with another source. Thus, the adductory, adductory plus F0, and adductory plus respiratory were considered one group with all other tremor types being considered members of the second group. Statistical analyses used in Experiment 1 were also applied to the computed variables in Experiment 2. Two discriminating variables emerged from the stepwise DFA: Peak Amplitude (Full signal), and Peak Frequency (4000 Hz band). The DFA classification function accurately classified 88.9% of the tremor samples as members of their designated group (with 88.9% on cross-validation, chance = 50%). The only misclassifications occurred for the isolated adductory and adductory plus respiratory tremors with extents of 10%. The two variables that were significant discriminators in this model were also included in the DFA model in the classification task of Experiment 1. The high degree of classification accuracy suggests that there is information specific to tremor source in the amplitude envelope spectra even in combination with additional tremor sources.

DISCUSSION

The purpose of the two experiments reported here was to explore the possibility that information about the source of a vocal tremor may reside in temporal regularities in the amplitude envelope for speech. If such information existed and was

TABLE 5.
Summary of the Parameters of Both the Synthesized Combination Tremors and Isolated Tremors for a Total of 18 Vocal Tremor Samples

Tremor Type	Extent of Modulation (%)	Rate of Modulation (Hz)	Steady-State Vowel	Fundamental Frequency (Hz)
Adductory + F0	2.5; 5; 10	5	[a]	100
Adductory + Resp	2.5; 5; 10	5	[a]	100
F0 + Resp	2.5; 5; 10	5	[a]	100
Adductory only*	2.5; 5; 10	5	[a]	100
F0 only*	2.5; 5; 10	5	[a]	100
Respiratory only*	2.5; 5; 10	5	[a]	100

Notes: Each tremor sample has a total duration of 3 seconds.
* Tremor samples from Exp. 1.

robust, it would provide not only an additional noninvasive tool for diagnosis of vocal tremor source but also provide a mapping of acoustic perturbations in speech output to their physiological source, which could be useful in therapy. To determine whether acoustic amplitude envelope measures were a viable means of distinguishing tremor source, we used a set of metrics that had previously been demonstrated to be useful for classifying speech disorder subtypes—the EMS metrics described by Liss et al.[40] We computed these metrics for a set of sustained vowels that were simulated to have three different types of vocal tremor with known sources. These acoustic measures were then entered as discriminating variables in a DFA to determine if the signals could be accurately grouped according to the source of the tremor used in creating the signal. Experiment 1 demonstrated that varying types of isolated vocal tremor could be accurately discriminated well above chance. Experiment 2 provided evidence that a tremor type can be accurately categorized even when paired with a second tremor source (when there is no presumption of physiological interaction). Together these experiments provide support for the utility of amplitude envelope spectra measurements for distinguishing tremor sources.

Several of the empirical design aspects of this study were motivated by the exploratory nature of the question being posed: is there information in the amplitude envelope of speech related to vocal tremor source? Whereas this design has provided fairly unambiguous positive evidence for the existence of such information, it remains to be determined just how robust this information is. We provide some cautionary notes about interpretation and directions for future research related to three design features of this study in the following: (i) the use of simulated tremor stimuli; (ii) the use of the specific EMS measures; and (iii) the use of stepwise discriminant analysis.

The use of a computational model to simulate the speech stimuli was motivated by the need to specify with certainty the source of the tremor (to assess the accuracy of categorization) and a desire to independently vary a number of parameters such as f0, tremor rate, tremor extent, etc. The resulting database provided a great deal of acoustic variability to challenge the categorization performed by the discriminant analysis, but also provided enough control to be able to examine what parameters affected categorization performance (eg, the modulation extent in Experiment 2). Whereas this database provides an initial evaluation of amplitude envelope metrics for determining tremor source, further testing will be required with additional sources of variability to better determine the robustness of such measures. For example, in the current experiments a single (male) vocal tract was used for all simulations (although the shape of the vocal tract varied to produce three different vowels in Experiment 1). Whereas the inclusion of vocal tract (and gender) variation will add extensive variability to the acoustics of the database, this variability is likely to be less evident in the temporal structures measured in the amplitude envelope. In fact, Carbonell et al[39] report that the EMS measures are not particularly adept at discriminating talker gender suggesting that varying gender would not have much systematic effect on amplitude envelope measures. Nevertheless, future tests should include variations in vocal tract anatomy and variations in the complexity of the stimuli from isolated vowels to words to phrases. There is some evidence that the perceptual effects of tremor differ across different phonetic segments and phrase types (eg, enhanced perception of vocal tremor for sustained vowels rather than connected speech).[1] It would be interesting to know whether the ability of amplitude envelope measures to detect and classify vocal tremor patterns occurs similarly to perception. One of the benefits of the current modeling approach is that these additional sources of variation can be added parametrically to the database to determine whether the EMS measure can accommodate these further complications as one moves closer to real cases of vocal tremor.

The EMS-dependent measures were the same as those used by Liss et al[40] to discriminate dysarthric subtypes. Whereas the previous success of these measures made them worthwhile starting points for the present study, it should be noted that those measures were designed to predict intelligibility and perceptual errors (which they do quite well for dysarthric speech.[42] As a result, several of the measures are based on the energy in the envelope around 4 Hz, which appears to be important for intelligibility.[30,31,43] It is highly unlikely that these same variables are optimal for distinguishing vocal tremor sources. The fact that these variables carried sufficient information to accurately categorize many of the stimuli in the database by source is evidence that tremor source has systematic effects on amplitude envelope spectra, but it is likely that these results underestimate the potential accuracy that could be attained with more suitable variables. The fundamental approach behind EMS—filtering the signal into frequency bands, calculating the power spectrum of each amplitude envelope, and computing measures on these spectra—appears promising. Future progress will depend on careful examination of the envelope spectra for different types of vocal tremor to develop dependent variables that provide as much discriminative power as possible.

One way to start the search for better variables is to look at which variables were most predictive in the experiments presented here. One must be careful not to rely too much on the specific variables that end up in the final model of a stepwise DFA because a variable may enter the model not because it is the best measure of the relevant information but because it is correlated with several variables that are good measures of relevant information. However, one can see from the analyses a couple of important generalizations. The first is that the discriminative information appears to reside in the envelope of the full signal and the lower frequency bands (especially the 125- and 500-Hz octave bands). The top six discriminating variables in Experiment 1 come from this subset as does the top variable from Experiment 2 (of two variables in the final model). The second is that the full signal peak amplitude variable was an important discriminating variable in both experiments (coming up as the top variable in Experiment 2 and the second variable entered into the model in Experiment 1). This variable provides a measure of how much the envelope for the full signal is dominated by a single modulation rate (the tremor rate for these isolated vowel stimuli). This measure was much lower for F0 tremor (highest for respiratory tremor). On the other hand, the F0 tremor has a much larger effect on the modulation in the 125-Hz band—as measured by the peak amplitude (125 Hz) variable. If one creates a new variable that computes the relative peak amplitude in the full signal versus the 125-Hz band, one can correctly classify 65.8% of the stimuli by source with this variable alone (chance = 33%). This is a better classification performance than obtained with any of the variables in the original EMS set. This simple example demonstrates the benefit of searching for more efficient variables. Once one has a set of such variables, one can develop and test particular models for classification with DFA without using the stepwise method (which can inflate accuracy rates).

CONCLUSIONS

The experiments presented here provide evidence that information about vocal tremor source is potentially available in the amplitude envelopes of the full signal and select frequency bands of the speech signal. Whether such information is robust enough to serve as a guide for diagnosis and therapy will require additional experiments with increased variability of signals and variables that are more specifically designed for the particular classification task.

Acknowledgments

This work was supported by a research grant from the National Institute on Deafness and Other Communication Disorders (Grant 5 R01 DC 4674) to the fourth author.

REFERENCES

1. Lederle A, Barkmeier-Kraemer J, Finnegan E. Perception of vocal tremor during sustained phonation compared with sentence context. *J Voice*. 2012;26:668.e1–668.e9.
2. Barkmeier-Kraemer J, Lato A, Wiley K. Development of a voice treatment program for management of a client with essential vocal tremor. *Semin Speech Lang*. 2011;32:43–57.
3. Bove M, Daamen N, Rosen C, Wang CC, Sulica L, Gartner-Schmidt J. Development and validation of the vocal tremor scoring system. *Laryngoscope*. 2006;116:1662–1667.
4. Brown JR, Simonson J. Organic voice tremor: a tremor of phonation. *Neurology*. 1963;13:520–525.
5. Dromey C, Warrick P, Irish J. The influence of pitch and loudness changes on the acoustics of vocal tremor. *J Speech Lang Hear Res*. 2002;45: 879–890.
6. Finnegan EM, Luschei ES, Gordon JD, Barkmeier JM, Hoffman HT. Increased stability of airflow following botulinum toxin injection. *Laryngoscope*. 1999;109:1300–1306.
7. Lester RA, Barkmeier-Kraemer J, Story BH. Physiologic and acoustic patterns of essential vocal tremor. *J Voice*. 2013;27:422–432.
8. Ludlow CL, Bassich CJ, Connor NP, Coulter DC. Phonatory characteristics of vocal fold tremor. *J Phon*. 1986;14:509–515.
9. Warrick P, Dromey C, Irish JC, Durkin L. The treatment of essential voice tremor with botulinum toxin A: a longitudinal case report. *J Voice*. 2000;14: 410–421.
10. Warrick P, Dromey C, Irish JC, Durkin L, Pakiam A, Lang A. Botulinum toxin for essential tremor of the voice with multiple anatomical sites of tremor: a crossover design study of unilateral versus bilateral injection. *Laryngoscope*. 2000;110:1366–1374.
11. Hachinski VC, Thomsen IV, Buch NH. The nature of primary vocal tremor. *Can J Neurol Sci*. 1975;2:195–197.
12. Koda J, Ludlow CL. An evaluation of laryngeal muscle activation in patients with voice tremor. *Otolaryngol Head Neck Surg*. 1992;107:684–696.
13. Tomoda H, Shibaski H, Kuroda Y, Shin T. Voice tremor: dysregulation of voluntary expiratory muscles. *Neurology*. 1987;31:117–122.
14. Adler CH, Bansberg SF, Hentz JG, et al. Botulinum toxin type A for treating voice tremor. *Arch Neurol*. 2004;61:1416–1420.
15. Brin MF, Blitzer A, Velickovic M. Movement disorders of the larynx. In: Blitzer A, Brin M, Ramig L, eds. *Neurologic Disorders of the Larynx*. New York, NY: Thieme Medical Publishers, Inc.; 1992:260.
16. Chien HF, Sanchez TG, Ubirajara Sennes L, Barbosa ER. Endonasal approach of salpingopharyngeus muscle for the treatment of ear click related to palatal tremor. *Parkinsonism Relat Disord*. 2007;13:254–256.
17. Gonzalez-Alegre P, Kelkar P, Rodnitzky RL. Isolated high-frequency jaw tremor relieved by botulinum toxin injections. *Mov Disord*. 2006;21: 1049–1050.
18. Krause E, Heinen F, Gurkov R. Difference in outcome of botulinum toxin treatment of essential palatal tremor in children and adults. *Am J Otolaryngol*. 2010;31:91–95.
19. Lou J, Jankovic J. Essential tremor: clinical correlates in 350 patients. *Neurology*. 1991;41:234–238.
20. Deuschl G, Bain P, Brin M, Ad Hoc Scientific Committee. Consensus statement of the Movement Disorder Society on tremor. *Mov Disord*. 1998; 13(suppl 3):2–23.
21. Koller WC, Busenbark K, Miner K. The relationship of essential tremor to other movement disorders: report on 678 patients. *Ann Neurol*. 1994;35: 717–723.
22. Sulica L, Louis ED. Clinical characteristics of essential voice tremor: a study of 34 cases. *Laryngoscope*. 2010;3:516–528.
23. Winholtz WS, Ramig LO. Vocal tremor analysis with the vocal demodulator. *J Speech Hear Res*. 1992;35:562–573.

24. Buder EH, Strand EA. Quantitative and graphic acoustic analysis of phonatory modulations: the modulogram. *J Speech Lang Hear Res*. 2003;46: 475–490.
25. Jiang J, Lin E, Wu J, Gener C, Hanson DG. Effects of simulated source of tremor on acoustic and airflow voice measures. *J Voice*. 2000;14: 47–57.
26. Lester RA, Story BH. Acoustic characteristics of simulated respiratory-induced vocal tremor. *Am J Speech Lang Pathol*. 2013;22:205–211.
27. Titze IR. Parameterization of the glottal area, glottal flow, and vocal fold contact area. *J Acoust Soc Am*. 1984;75:570–580.
28. Titze IR. *The Myoelastic Aerodynamic Theory of Phonation*. Iowa City, IA: National Center for Voice and Speech; 2006.
29. Story BH. A parametric model of the vocal tract area function for vowel and consonant simulation. *J Acoust Soc Am*. 2005;117:3231–3254.
30. Drullman R, Festen JM, Plomp R. Effect of reducing slow temporal modulations on speech reception. *J Acoust Soc Am*. 1994;95:2670–2680.
31. Drullman R, Festen JM, Plomp R. Effect of temporal envelope smearing on speech reception. *J Acoust Soc Am*. 1994;95:1053–1064.
32. Ghitza O. On the role of theta-driven syllabic parsing in decoding speech: intelligibility of speech with a manipulated modulation spectrum. *Front Psychol*. 2012;3:238.
33. Ghitza O. The theta-syllable: a unit of speech information defined by cortical function. *Front Psychol*. 2013;4:138.
34. Peelle JE, Davis MH. Neural oscillations carry speech rhythm through to comprehension. *Front Psychol*. 2012;3:320.
35. Peelle JE, Gross J, Davis MH. Phase-locked responses to speech in human auditory cortex are enhanced during comprehension. *Neuroscience*. 2013; 23:1378–1387.
36. Luo H, Poeppel D. Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron*. 2007;54:1001–1010.
37. Carbonell KM, Brenner D, Lotto AJ. "Discriminating languages with general measures of temporal regularity and spectral variance." 164th Meeting of the Acoustical Society of America; October 24, 2012.
38. LeGendre SJ, Liss JM, Lotto AJ, Utianski R. "Talker recognition using envelope modulation spectra." 158th Meeting of the Acoustical Society of America, San Antonio, TX; October 26, 2009.
39. Carbonell KM, Lansford K, Utianski R, Liss JM, Lotto AJ. "Discriminating language and talker using non-linguistic measures of rhythm, spectral energy and f0." 162nd Meeting of the Acoustical Society of America; November 4, 2011.
40. Liss J, Legendre S, Lotto A. Discriminating dysarthria type from envelope modulation spectra. *J Speech Lang Hear Res*. 2010;53:1246–1255.
41. Story BH. Phrase-level speech simulation with an airway modulation model of speech production. *Comput Speech Lang*. 2013;27:989–1010.
42. LeGendre SJ, Liss JM, Lotto AJ. "Discriminating dysarthria type and predicting intelligibility from amplitude modulation spectra." 157th Meeting of the Acoustical Society of America; May 18, 2009.
43. Houtgast T, Steeneken JM. A review of the mtf concept in room acoustics and its use for estimating speech intelligibility in auditoria. *J Acoust Soc Am*. 1985;77:1069–1077.