The effects of physiological adjustments on the perceptual and acoustical characteristics of simulated laryngeal vocal tremor

Rosemary A. Lester^{a)} and Brad H. Story

Department of Speech, Language, and Hearing Sciences, University of Arizona, Tucson, Arizona 85721, USA

(Received 3 October 2014; revised 17 July 2015; accepted 17 July 2015; published online 18 August 2015)

The purpose of this study was to determine if adjustments to the voice source [i.e., fundamental frequency (F0), degree of vocal fold adduction] or vocal tract filter (i.e., vocal tract shape for vowels) reduce the perception of simulated laryngeal vocal tremor and to determine if listener perception could be explained by characteristics of the acoustical modulations. This research was carried out using a computational model of speech production that allowed for precise control and manipulation of the glottal and vocal tract configurations. Forty-two healthy adults participated in a perceptual study involving pair-comparisons of the magnitude of “shakiness” with simulated samples of laryngeal vocal tremor. Results revealed that listeners perceived a higher magnitude of voice modulation when simulated samples had a higher mean F0, greater degree of vocal fold adduction, and vocal tract shape for /i/ vs /a/. However, the effect of F0 was significant only when glottal noise was not present in the acoustic signal. Acoustical analyses were performed with the simulated samples to determine the features that affected listeners’ judgments. Based on regression analyses, listeners’ judgments were predicted to some extent by modulation information present in both low and high frequency bands. © 2015 Acoustical Society of America.

[<http://dx.doi.org/10.1121/1.4927561>]

[ZZZ]

Pages: 953–963

I. INTRODUCTION

Vocal tremor is a voice disorder that affects individuals with neurological impairments including essential tremor, Parkinson disease, cerebellar dysfunction, and dystonia (Deuschl *et al.*, 1998; Koller *et al.*, 1994; Lou and Jankovic, 1991). Of the 10 million Americans diagnosed with *essential* tremor, the most common movement disorder associated with tremor, it is estimated that 18%–30% develop *essential* vocal tremor (Koller *et al.*, 1994; Lou and Jankovic, 1991; Louis, 2005). Current approaches for medical management of essential vocal tremor have inconsistent effects on voice production, and evidence to support behavioral management is limited (Barkmeier-Kraemer *et al.*, 2011; Koller *et al.*, 1987; Koller *et al.*, 1985; Massey and Paulson, 1985; Warrick *et al.*, 2000a). The challenges in managing essential vocal tremor stem from the complex relations among the physiological, acoustical, and perceptual aspects of the disorder. It is important to understand these relations to develop treatment methods that improve voice production for individuals with essential vocal tremor and minimize their communication impairment. The purpose of this study was to: (1) investigate the effects of physiological adjustments on listeners’ perception of vocal tremor to understand potential compensatory speaking strategies and (2) determine if listener perception could be explained by characteristics of the acoustical modulations. It was hypothesized that perception of the magnitude of vocal tremor would be affected by modulation information in high frequency regions of the

voice spectrum, which could be altered by physiological adjustments to the voice.

Essential vocal tremor involves atypical modulation of the frequency and intensity of the voice (Brown and Simonson, 1963; Ludlow *et al.*, 1986). This disorder is associated with abnormal neurological functioning that causes nearly periodic contraction of muscles in the vocal system. Within the larynx, these muscle contractions may change the length of the vocal folds and, as a result, modulate the fundamental frequency (F0) of vibration. As the F0 modulates, the harmonics also modulate. This is demonstrated in Fig. 1 where the amplitude-frequency trajectory of the first 30 harmonics is plotted over the time course of a simulated vowel with a 10% extent of F0 modulation. It is noted that even this small percentage change in F0 can generate a wide range of frequency modulation in the higher harmonics due to the proportional nature of the shift, although the percentage of change is the same. Simulated changes in vocal fold length with modulation of the F0 also produce co-modulation of the intensity of the acoustic output (Lester *et al.*, 2013). This is consistent with the resonance-harmonics interaction described by Horii and Hata (1988), wherein modulation of the F0 produces modulation of the intensity as the harmonics sweep in and out of the vocal tract resonances.

Adjustments to the voice have been suggested to be beneficial in reducing listener perception of vocal tremor. It is hypothesized that the effects of some of these adjustments are related to changes in the amplitude of the higher harmonics. Barkmeier-Kraemer *et al.* (2011) noted that, in a case study of one individual with essential vocal tremor, the perception of vocal tremor severity was reduced when the individual produced a breathy voice. This may be related to

^{a)}Current address: Department of Otorhinolaryngology, Mayo Clinic, Phoenix, AZ 85054, USA. Electronic mail: Lester.Rosemary@mayo.edu

FIG. 1. (Color online) Spectrum of simulated modulation of F0. The mean F0 is 130 Hz, the modulation rate is 5 Hz, and the modulation extent is 10% (i.e., 13 Hz extent; F0 range of 117–143 Hz). Each point represents the amplitude of an individual harmonic on the frequency scale expressed in Hz.

the fact that voice produced with incomplete adduction of the vocal folds (associated with a breathy vocal quality) involves lower amplitude of the higher harmonics as compared with a normal or pressed voice produced with greater adduction of the vocal folds (Titze, 1994). Barkmeier-Kraemer *et al.* (2011) also reported an effect of pitch manipulation on the perception of severity of vocal tremor. This observation may be associated with differences in the amplitude of the higher harmonics for different fundamental frequencies. That is, independent of the vocal tract filter, the harmonics decrease in amplitude by about 12 dB per octave (Titze, 1994). Therefore, the amplitude of the harmonics in any given portion of the voice spectrum would be lower when the F0 is low in frequency than when the F0 is high in frequency. For example, if the F0 were 100 Hz, the amplitude of the harmonic at 800 Hz would be 36 dB less than the amplitude of the F0; however, if the F0 were 200 Hz, the amplitude of the harmonic at 800 Hz would be 24 dB less than the amplitude of the F0.

Vocal tract configurations for vowels are also known to affect the amplitude of higher harmonics and could affect the perception of magnitude of voice modulation (Fant, 1960). For example, the vowel /i/ typically has widely spaced first and second formants, whereas /a/ typically has more closely spaced first and second formants in the lower frequency region (Peterson and Barney, 1952). As such, it is expected that modifications to the vocal tract configuration could attenuate the amplitude of the higher harmonics and reduce the perception of voice modulation. The usefulness of these manipulations to the voice source and vocal tract filter for reducing the perception of modulation of voice requires further investigation.

The aim of study 1 was to use computer-simulated voice samples involving primary modulation of the F0 to determine if adjustments to the voice source (i.e., vocal fold length, degree of vocal fold adduction) and vocal tract filter (i.e., vocal tract configuration for vowels) reduce the perception of the magnitude of voice modulation. It was hypothesized that listeners would perceive a higher magnitude of voice modulation when simulated samples had a higher

mean F0 corresponding to an increased vocal fold length, greater degree of vocal fold adduction corresponding to a pressed voice, or a vocal tract shape with clustering of resonant peaks in higher frequency regions (i.e., /i/). The aim of study 2 was to determine the acoustical characteristics that were associated with differences in perception of the magnitude of voice modulation. It was hypothesized that the extent of modulation in the higher frequency regions and/or the mean amplitude in the higher frequency regions would predict listeners' judgements of the magnitude of modulation of voice. The results of these studies are expected to be useful in identifying adjustments to the voice source and vocal tract filter that might be produced by speakers to reduce the perception of voice modulation, as well as to understand the acoustical bases for the perception of the magnitude of modulation of voice.

II. STUDY 1

A. Method

1. Computational model

Simulations of vocal tremor were produced using a kinematic model of the vocal folds (Titze, 1984, 2006) coupled to a wave-reflection model of the trachea and vocal tract (Liljencrants, 1985; Story, 1995, 2005) using custom-written MATLAB (The Mathworks, 2011, Version 7.13.0.564 [R2011b]) scripts and compiled c code. These models allowed for control of the pressure at the level of the bronchi (representing tracheal pressure), the resting length and thickness of the vocal folds, the fundamental frequency of vibration (F0), the degree of adduction of the vocal folds, and the shape of the vocal tract. The propagation of acoustic waves in the trachea and vocal tract accounted for energy loss due to yielding walls, viscosity, heat conduction, and radiation.

The kinematic model represented the medial surfaces of the vocal folds. Vibration of these surfaces generated a time-varying glottal area, which converted dc airflow from the trachea into a periodic series of pulses. The position and vibration of the vocal fold surfaces were defined by a combination of user-input and rules derived from anatomical and physiological studies (Titze, 2006). The structure and function of the model is the same as described in Samlan and Story (2011). For this study, the F0 characteristics were specified in the model by the mean F0, F0 modulation rate, and F0 modulation extent (as a percentage of the mean F0). This forced-modulation of the F0 produced changes in the length and vibratory pattern of the vocal folds and, as a result, changes in the glottal area. That is, because the model was driven at a specified F0, the instantaneous vocal fold length was determined by the F0. This is unlike a natural self-oscillating system in which the vocal fold length, in part, determines the F0. Because this is a kinematic model, the mass and stiffness of the vocal folds were not represented. Parameter settings that were specified by the user are listed in Table I and discussed below. All other parameters were determined by rules as described by Titze (2006).

TABLE I. Parameter settings for the computational model used to simulate vocal tremor.

Parameter	Setting	Parameter	Setting
Fundamental frequency (F0)	110, 123, 165, 220, 247, 262 Hz	Starting phase bilaterally	0 radians
F0 modulation rate	5 Hz	Pressure	8000 dyn/cm ² ; 8 cm H ₂ O
F0 modulation extent	1, 5, 10, 20%	Epilaryngeal area	0.5 cm
Superior adduction (ξ_{02}) bilaterally	0, 0.1, 0.2 cm	Length of each section in vocal tract area function	0.35 (female) 0.40 (male)
Surface bulging (ξ_b) bilaterally	0.07 (female) 0.1 (male)	Vowel shapes	/ə, i, a/
Nodal point ratio (z_n) bilaterally	0.7	Time	1 s

Two sets of stimuli were generated using this computational model. One set of stimuli included glottal noise and was used for study 1a; the other set of stimuli did not include glottal noise and was used for study 1b. With the first set of stimuli, the breathy vocal quality was produced by incomplete glottal closure, which produced noise due to turbulence within the glottis. This was random broadband noise, with amplitudes ranging from -0.5 to 0.5 and frequencies ranging from 300 to 3000 Hz. The noise was added to the glottal flow when a threshold value was exceeded for the Reynolds number, which was calculated using the instantaneous glottal flow, length of the glottis, air density, and air viscosity. As such, the Reynolds number modulated as the vocal fold length modulated; therefore, the noise modulated with vocal fold length modulations. Details concerning these calculations are described in [Lester \(2014\)](#) and [Samlan and Story \(2011\)](#). In the second set of stimuli, this frication noise was not included in the signal. These simulations were used to test that, if there were significant findings in study 1a, the findings were not due to the presence of modulated noise in the signal.

The pressure changes associated with variations in glottal area interacted with the vocal tract pressures to produce the glottal flow. The acoustic energy produced by the glottal flow was filtered by the trachea and the vocal tract. As a result, the characteristics of the acoustic output were influenced by the resonances of both the trachea and the configuration of the vocal tract. Vocal tract area functions known to produce target vowels were used to determine vocal tract shapes ([Bunton and Story, 2010](#)). Two versions of the model were used for this study: one that was based on the airway features of a female and another based on the airway features of a male. Schematic representations of these models are shown in Fig. 2, where, for purposes of demonstration, the female vocal tract is configured as an /a/ vowel and the male as an /i/ vowel.

2. Audio stimuli

The female and male versions of the model were used to simulate productions of sustained vowels using vocal tract shapes representative of the vowels /ə, i, a/. The vocal tract shapes for the three vowels were based on [Story \(2005\)](#) for the female and on [Story et al. \(1996\)](#) for the male. The resonance frequencies of each vocal tract shape were calculated directly with a transmission line (frequency domain) method and yielded the values shown in Table II. Note that these

values serve as an idealized reference in that there is no interaction with the voice source or coupling to the trachea. In a strict sense, they are resonance frequencies rather than formants, but for simplicity are denoted in the table with the typical formant notation F1, F2, etc.

Each vowel was simulated with a duration of 1 s. For each stimulus, a 0.1 s long silence was included before and after each vowel, and a 0.05 s fade in and out was imposed for each vowel to avoid a click at the onset and offset. The sampling frequency was 44 100 Hz. The F0 values for these simulations were 110, 123, and 165 Hz for the male model and 220, 247, and 262 Hz for the female model. These frequencies are within the average range for adult males and females and also correspond to musical notes to be used in future studies with humans. The rate of F0 modulation was 5 Hz, which is within the range of rates typically exhibited by individuals with vocal tremor ([Gamboa et al., 1998](#)). The extents of F0 modulation were 1%, 5%, 10%, and 20%, which are also within the range of extents that are typically exhibited in this population ([Dromey et al., 2002](#); [Ramig and Shipp, 1987](#)). The degrees of vocal fold adduction were set to 0, 0.1, and 0.2 cm to represent slightly pressed, moderate, and low degrees of vocal fold adduction, respectively. The differences in the glottal area and glottal flow between these three degrees of vocal fold adduction are

FIG. 2. (Color online) Wave-reflection model of the trachea and vocal tract. These are based on the airway features of a female producing /a/ (left) and male producing /i/ (right). P_L represents the lung pressure. The glottal area, glottal flow, and output sound pressure waveforms indicated in their respective locations within the model.

TABLE II. Resonant frequencies for simulated vowels /ə, i, a/ using the female and male versions of the computational model.

Model	Vowel	F1	F2	F3	F4
Female	/ə/	684	1674	2875	4201
	/i/	358	2871	4538	5442
	/a/	778	1315	3045	4279
Male	/ə/	601	1504	2492	3653
	/i/	312	2552	3921	4925
	/a/	677	1093	2724	3586

shown in Fig. 3. The pressure supplied to the vocal folds was set to 8000 dyn/cm², which corresponds to a normal pressure for speech production of approximately 8 cm H₂O. Further details related to the model parameters can be found in Story (2013).

3. Listeners

Twenty-one listeners (17 female, 4 male) with a mean age of 23 years (range = 19–58 years) were included in study 1a, and 21 listeners (18 female, 3 male) with a mean age of 22 years (range = 19–39 years) were included in study 1b. Prior to the experimental trials, all listeners passed a bilateral hearing screening using pulsed pure tones at 500, 1000, 2000, and 4000 Hz at 25 dB (American Speech-Language-Hearing Association, 1997). All listeners reported and exhibited a good understanding of written and spoken English and denied a history of cognitive disorders. This study was approved by the University of Arizona Institutional Review Board (UA IRB).

4. Listening task

Listeners were seated individually in a sound-treated booth. Audio stimuli were presented with an intensity of 60 dB sound pressure level over a loudspeaker placed 1 m in front of the listener. Stimuli were presented and listeners’ responses were recorded using the ALVIN interface (Hillenbrand and Gayvert, 2005). The inter-stimulus interval was set to 0.5 s and the inter-trial interval was set to 1 s. With the 0.1 s-long silence included at the beginning and end of each stimulus, there was a 0.7 s silence between each

vowel within a pair and a 1.1 s silence between the second vowel in one trial and the first vowel in the next trial. The trial advanced as soon as listeners made their selection of which vowel in the pair was “shakier.”

Prior to the experiment, each listener participated in a brief training protocol consisting of 16 pairs of simulated vowels with different extents of modulation (1%, 5%, 10%, 20%), one F0 (either 123 Hz for the male vocal tract or 247 Hz for the female vocal tract), one degree of vocal fold adduction (corresponding to a normal vocal quality), and one neutral vowel (/ə/). Each pair was presented in AB and BA orders. A pair-comparison approach was selected over using a perceptual rating scale because pair-comparisons were successfully used in a previous study of the perception of respiratory-induced vocal tremor and had high intra-rater agreement (Farinella *et al.*, 2006). Listeners were instructed to identify which stimulus in each pair was shakier. Listeners were allowed to replay the pairs of stimuli as many times as needed to make their judgments and to take as many rest breaks as needed. They were instructed to “guess” if they were unsure which item in the pair was shakier.

“Shaky” was selected as the descriptor for these studies because previous studies demonstrated that 70/70 naive listeners independently described sustained vowels produced by individuals with vocal tremor as shaky, and this descriptor has been used in previous perceptual studies of vocal tremor (Lederle *et al.*, 2012). In addition, an unpublished pilot study was conducted to determine if listeners could perceive differences in the magnitude of shakiness using pair comparison judgments. The stimuli for this study were simulated productions of /i/ with an F0 of 267 Hz, a modulation rate of 5 Hz, and F0 modulation extents of 1%, 5%, 10%, and 20%. Listeners were presented with pairs of stimuli with different extents of modulation and were asked to determine which sample in each pair was shakier. Listeners performed this task with a mean accuracy of 92.5% across five listeners. That is, in the vast majority of cases, the sample with the highest extent of F0 modulation was selected as the shakier stimulus in the pair. The range of accuracy was 94%–100% for four listeners and 75% for one listener. Intra-rater reliability was determined through repeated presentations of 75% of the stimuli and was found to be 90%. The results of these studies indicated that listeners can detect differences in

FIG. 3. Glottal area (left panel) and glottal flow (right panel) waveforms representing high (lower trace), moderate (middle trace), and low (upper trace) degrees of vocal fold adduction.

shakiness between paired stimuli with high accuracy and reliability.

Additional trials were conducted to determine if listeners could identify when there was no difference in the magnitude of shakiness. Four listeners were presented with pairs of vowels that were identical and pairs of vowels differing by one feature as described below. They were given the option of choosing that the first vowel in the pair was shakier, the second vowel in the pair was shakier, or the first and second vowels in the pair were equally shaky. They identified the identical pairs with a mean accuracy of 95% (range of 91%–100%). These trials indicated that participants can identify stimuli that are identical when given the choice. As such, for the training and experimental trials, participants were presented with only two choices: the first vowel is shakier or the second vowel is shakier.

For the experimental trials, participants were again presented with pairs of stimuli and were asked to identify which stimulus in the pair was shakier. The stimuli in these pairs were either identical or differed by one characteristic only (i.e., F0, degree of vocal fold adduction, or vocal tract shape). For example, one stimulus in a pair had an F0 of 110 Hz, a low degree of vocal fold adduction (corresponding to a breathy voice), /i/ vocal tract shape, and 1% modulation; the other stimulus in the pair had an F0 of 165 Hz, a low degree of vocal fold adduction, /i/ vocal tract shape, and 1% F0 modulation. With all possible combinations of F0, degree of vocal fold adduction, vocal tract shape, and extent of F0 modulation for the female vocal tract or the male vocal tract, counterbalancing for order using AB and BA presentations, and identical pairs, 256 trials were presented to each listener. Listeners were again given the option of replaying the stimuli and taking rest breaks at their discretion. They were also instructed to “guess” if they were unsure which item in the pair was shakier. Presentation of the female and male simulations was on two separate days for each listener; the order of presentation was randomly assigned and counterbalanced. The procedures for studies 1a and 1b were identical, except the stimuli in study 1b did not include glottal noise.

5. Data analysis

For each contrast (F0, degree of vocal fold adduction, vocal tract shape), the data were analyzed to determine the percentage of trials that the participants’ responses were consistent with the hypotheses [i.e., the percentage of trials when high F0 was selected over low F0, high degree of vocal fold adduction (corresponding to a pressed vocal quality) was selected over a low degree of vocal fold adduction (corresponding to a breathy vocal quality), and a vocal tract shape representing /i/ was selected over a vocal tract shape representing /a/ as being shakier]. This was determined separately for the glottal noise and no glottal noise conditions. Intra-rater reliability was calculated by determining the percentage of trials (out of 4) when each participant’s judgments were consistent across items. This analysis was performed in SPSS (IBM, version 22) using the aggregate function with the break variables of participant and stimulus pair. Inter-rater agreement was calculated by determining

the percentage of trials when all participants’ judgments were the same across items. This analysis was performed using an aggregate function with the break variable of stimulus pair only.

6. Reliability and agreement

Intra-rater reliability was 75% for the female stimuli with glottal noise, 75% for the male stimuli with glottal noise, 71% for the female stimuli without glottal noise, and 73% for the male stimuli without glottal noise. The inter-rater agreement was 64% for the female stimuli with glottal noise, 64% for the male stimuli with glottal noise, 59% for the female stimuli without glottal noise, and 59% for the male stimuli without glottal noise.

7. Statistical analyses

The data were analyzed using a repeated measures analysis of variance (ANOVA) with glottal noise as the between-subjects factor and with sex of the computational model (male, female) and contrast (F0, degree of vocal fold adduction corresponding to the vocal quality, vocal tract shape corresponding to the vowel) as the within-subjects factors in SPSS (IBM, 2013, version 22). Data were collapsed across extent of modulation because initial analyses did not reveal a pattern in participants’ judgments based on extent. Planned one-tailed, one-sample t-tests with a test value of 50 (to account for chance) were performed in SPSS. *Post hoc* analyses of effect size were calculated using Cohen’s *d* with a mean *H*₀ of 50, mean of the actual data, and standard deviation of the actual data in G*Power (Faul *et al.*, 2007, version 3.1.7).

B. Results

The repeated measures ANOVA tests of within-subjects effects revealed that there was a significant main effect of contrast [$F(2,80) = 10.790$, $p < 0.001$] when a Greenhouse-Geisser correction was applied due to violation of the assumption of sphericity, as indicated by a significant result of Mauchley’s Test of Sphericity ($p = 0.005$). There was a significant interaction of contrast and glottal noise [$F(2,80) = 6.124$, $p = 0.006$] when a Greenhouse-Geisser correction was applied. Tests of between-subjects effects revealed that there was a significant intercept [$F(1,40) = 61.183$, $p < 0.001$]. The data are presented in Fig. 4.

T-tests with the data collapsed across extent and sex of the model revealed a significant effect of degree of vocal fold adduction [$(M = 75.11, SD = 13.505)$; $t(41) = 12.051$, $p < 0.001$, $d = 1.859$] and vocal tract shape [$(M = 58.30, SD = 16.225)$; $t(41) = 3.314$, $p = 0.001$, $d = 0.512$] with glottal noise and a significant effect of F0 [$(M = 58.22, SD = 12.113)$; $t(41) = 4.399$, $p < 0.001$, $d = 0.678$], degree of vocal fold adduction [$(M = 60.68, SD = 16.449)$; $t(41) = 4.207$, $p < 0.001$, $d = 0.649$], and vocal tract shape [$(M = 56.92, SD = 9.490)$; $t(41) = 4.726$, $p < 0.001$, $d = 0.421$] without glottal noise.

FIG. 4. (Color online) Percentage of responses consistent with hypotheses for male and female simulations with contrasts of F0, degree of vocal fold adduction corresponding to the vocal quality, and vocal tract shape corresponding to the vowel for the conditions: with glottal noise (left) and without glottal noise (right).

C. Discussion

The results of this study revealed that the degree of vocal fold adduction and vocal tract shape affected listeners' perception of the magnitude of modulation of voice when glottal noise was present in the acoustic signal (as in the human voice when turbulence is generated due to incomplete glottal closure). A greater degree of vocal fold adduction corresponding to a pressed voice was perceived as being shakier than a lower degree of vocal fold adduction corresponding to a breathy voice. Based on analyses of effect sizes, this was a large effect (Cohen, 1969). The vocal tract shape for /i/, which is has clustering of resonant peaks in higher frequency regions, was perceived as shakier than the vocal tract shape for /a/, which enhances the amplitude of the lower harmonics. Effect size analyses indicated that this was a moderate effect.

When glottal noise was not present in the acoustic signal, the mean F0 representing vocal fold length, the degree of vocal fold adduction, and the vocal tract shape affected listeners' perception of the magnitude of modulation of voice. That is, a higher F0, a greater degree of vocal fold adduction corresponding to a pressed voice, and a vocal tract shape for /i/ were perceived as being shakier. The corresponding effect sizes were moderate. Although it is not typical for a human to produce a breathy voice without glottal noise as it was in the computational model, the simulations without glottal noise were useful in revealing that F0 significantly affected perceived shakiness of the voice only when glottal noise was not present in the signal. This finding may indicate that the glottal noise, which was in the 300–3000 Hz frequency region, masked small differences in the amplitude of the higher harmonics between the high and low fundamental frequencies.

Intra-rater reliability results suggest that listeners were consistent in their judgments of the magnitude of modulation of voice; they made the same judgements for approximately three out of four presentations of each stimulus. However, inter-rater agreement was lower. This suggests that listeners may have different strategies for judging the magnitude of modulation of voice. This also may suggest that there were some differences between listeners that were not accounted for like musical training.

It is presumed that these perceptual results are related to a reduction in the amplitude of the higher harmonics with the simulated manipulations to the voice source and vocal tract filter. However, acoustical analyses are needed to determine the acoustical characteristics that account for listeners' perceptual responses. It is hypothesized that the mean

amplitude and/or extent of amplitude modulation in high frequency bands predict listeners' judgments.

III. STUDY 2

A. Method

1. Acoustical and perceptual data

The acoustic signals for study 2 were the experimental audio stimuli used in study 1. All of the stimulus pairs from studies 1a and 1b, except for the identical pairs, were included in the acoustical analyses for study 2.

2. Acoustical analyses

Acoustical analyses were performed using custom-written MATLAB (The Mathworks, 2011, version 7.13.0.564 [R2011b]) scripts based on the American National Standard Specification for Octave-Band and Fractional-Octave-Band Analog and Digital Filters [ANSI S1.1-1986 (ASA 65-1986), (ANSI, 1993)]. The root mean square (rms) amplitudes within each octave band were normalized to the total rms amplitude (on a linear scale) for each full signal. The acoustic signals were then analyzed to determine the mean amplitude and the standard deviation of the amplitude within the six standard octave bands (center frequencies: 125, 250, 500, 1000, 2000, and 4000 Hz) shown in Fig. 5. The coefficient of variation of the amplitude was then calculated by dividing the standard deviation by the mean for each octave band to represent the extent of amplitude modulation within each band. Coefficient of variation was used rather than a direct measurement of extent because of the non-sinusoidal nature of the amplitude output for the octave bands, particularly in the higher frequencies. This was likely a result of harmonic energy modulating in and out of adjacent bands as the F0 modulated.

3. Statistical analyses

For each contrast (F0, degree of vocal fold adduction, vocal tract shape), the participants' perceptual judgments (i.e., stimulus 1 is shakier or stimulus 2 is shakier) were modeled using binary logistic stepwise regression with potential predictor variables representing (1) the ratios of the mean amplitudes in each of the six octave bands for each of the stimulus pairs and (2) the ratios of the coefficients of variation of amplitude in each of the six octave bands for each of the stimulus pairs. These two models were calculated separately for the glottal noise and no glottal noise conditions,

FIG. 5. (Color online) Six octave bands (with center frequencies of 125, 250, 500, 1000, 2000, 4000 Hz) used for acoustical analyses of simulated vowels with F0 modulation.

but not for the female and male stimuli to more precisely model the perceptual results in study 1.

The order of entry of the variables into the forward conditional stepwise regression model was based on a conditional likelihood ratio test with criterion of variable entry of $p < 0.05$ and removal of $p > 0.10$. That is, the variables that were entered into the model were the best predictors of the odds that a listener would choose one stimulus over another as being shakier. These models were calculated using SPSS (IBM, version 22).

B. Results

Examples of the results of the acoustical analyses are presented in Figs. 6 and 7. All acoustical analyses are presented in Lester (2014). The results of the regression analyses are presented in Table III.

C. Discussion

The results of this study indicate that low frequency bands containing the F0 as well as high frequency bands containing the higher harmonics may be used by listeners when making judgments about the magnitude of modulation of voice. However, the frequency regions that best predicted participants' responses depended on the contrasts within the stimulus pairs and the acoustical features of the stimuli in the pairs. Combinations of the mean amplitude and coefficient of variation of amplitude in the low and high frequencies were used for most contrasts. Although there were patterns in the variables used in the regression models for the F0, degree of vocal fold adduction, and vocal tract shape contrasts, the goodness of fit of all of the models was only fair based on the R-square values.

The limited goodness of fit of all of the models suggests that there are other acoustical features that might contribute to listeners' perception of modulation of voice besides the mean amplitude of low and high frequency bands and the extent of amplitude modulation within these bands, or that there are interactions between acoustical features that affect listeners' perception. It also suggests that listeners may have different perceptual strategies for judging the magnitude of modulation of voice. Kreiman and Gerratt (2012) found that there are interactions between the shape of the spectrum and the levels of harmonic energy and noise that affect listeners' perception of voice. Spectral shape and noise-to-harmonics ratio were not investigated in the current study, but it is possible that these interactions affected the results of the current study. It is also possible that the results were affected by noise masking small acoustical differences in the high

frequency regions or the lack of high frequency energy in the stimuli with a low degree of vocal fold adduction without glottal noise.

Two other perceptual phenomena are important to consider when interpreting the results of this study: the resonance-harmonic interaction (Hori and Hata, 1988) and auditory masking (Wegel and Lane, 1924). It is possible that the frequency bands used by listeners to make judgments about the shakiness of the voice were bands wherein harmonics fell in close proximity to vocal tract resonances. In these regions, there would be greater co-modulation of the amplitude, which may have contributed to the perception of shakiness. An increase in the amplitude of the harmonics in these frequency regions could also contribute to masking of weak harmonics by the strengthened harmonics. With the upward spread of masking, this may have reduced the likelihood that higher frequency bands would have been used to make judgments about the shakiness of the voice. These factors require further investigation.

D. Clinical applications

In this study, vocal quality had a large effect on participants' judgments of the magnitude of modulation of voice and had the highest mean number of responses consistent with the study hypotheses (i.e., a pressed voice would be judged as shakier than a breathy voice) compared with the other two contrasts (i.e., F0 and vowel) when glottal noise was included in the signal. Acoustical analyses revealed that breathy stimuli contained less energy above 1000 Hz for /a/ and above 2500 Hz for /i/ with both male and female signals, as compared with pressed stimuli. Additionally, in most cases, the mean amplitude of the low frequency bands containing the F0 were the highest for the breathy stimuli, and the mean amplitude of the high frequency bands were the highest for the pressed stimuli. Regression analyses revealed that the best predictors of participants' judgments with the breathy vs pressed stimuli were the mean amplitude and the coefficient of variation within both the low and high frequency bands.

These findings indicate that the production of a breathy voice might be a useful compensatory strategy for reducing the perception of vocal tremor in individuals with F0 modulation related to tremor affecting vocal fold length. The benefit of this manipulation of vocal quality appears to have acoustical rather than physiological underpinnings, based on these studies. In contrast, the success of manipulation of the F0 found by Barkmeier-Kraemer *et al.* (2011) may have been due to a physiological effect that did not occur in these simulations to drastically reduce listener

FIG. 6. (Color online) Narrowband spectrogram (left panel), octave band analyses (middle panel), and average spectrum (right panel) for the female /i/ with an F0 of 220 Hz, an F0 modulation extent of 5%, and a low degree of vocal fold adduction (corresponding to a breathy vocal quality) with glottal noise (upper row) and without glottal noise (middle row), and a high degree of vocal fold adduction (corresponding to a pressed vocal quality) (lower row). Figure titles indicate the sex of the model (fem/male), extent of F0 modulation in percent, vowel, F0, and breathy (b) vs pressed (p) vocal qualities.

perception of modulation of voice. Although the effect of F0 was significant with stimuli that did not contain glottal noise in this study, this does not seem to have clinical relevance because the result was only 8% above chance and because glottal noise is often present in the human voice. However, in a study using a three-mass model of the vocal folds, an adjustment to F0, which involved an alteration in the balance of activation of the cricothyroid and thyroarytenoid muscles, changed the extent of F0 modulation in the acoustic signal (Lester and Story, 2013). Therefore, it is possible that there would be an effect of F0 on listeners' perception of the magnitude of modulation of voice if simulations were generated using a different model or with the human voice.

Although changes in vocal tract shape for vowel production had only a moderate effect on listeners' perception of magnitude of modulation of voice (with listener judgements being only 8% above chance with glottal noise and 7% above chance without glottal noise) and would not likely be an easily implemented compensatory strategy for individuals with vocal tremor, it might be possible to modify an individual's overall vocal tract shape to reduce the amplitude of the higher harmonics. For example, singers have the ability to change the overall shape of the vocal tract to enhance the energy in around 3000 Hz while preserving the target vowel identity; this is known as the singer's formant (Sundberg, 1974; Weiss *et al.*, 2001). Therefore, it might be possible to train individuals with vocal tremor change the overall vocal

FIG. 7. (Color online) Narrowband spectrogram (left panel), octave band analyses (middle panel), and average spectrum (right panel) for the male /a/ with an F0 of 110 Hz, an F0 modulation extent of 10%, and a low degree of vocal fold adduction (corresponding to a breathy vocal quality) with glottal noise (upper row) and without glottal noise (middle row), and a high degree of vocal fold adduction (corresponding to a pressed vocal quality) (lower row). Figure titles indicate the sex of the model (fem/male), extent of F0 modulation in percent, vowel, F0, and breathy (b) vs pressed (p) vocal qualities.

tract shape to reduce the energy in the higher frequencies. However, depending on the speaker's F0 and the spoken vowel, it is possible that this adjustment could lead to reduced intelligibility (Hollien *et al.*, 2000). This requires further investigation.

The findings of this study are useful in considering the effectiveness of compensatory strategies for vocal tremor, and might also explain why laryngeal botulinum toxin injections are beneficial for some individuals with laryngeal vocal tremor. These injections typically result in decreased muscle activation levels (Burgin *et al.*, 1949; Ludlow *et al.*, 1990), which leads to a breathy vocal quality (Adler *et al.*, 2004; Warrick *et al.*, 2000a). This medical treatment of laryngeal

vocal tremor has been found to be beneficial in reducing patient perception of vocal tremor severity and functional disability (Adler *et al.*, 2004; Warrick *et al.*, 2000a; Warrick *et al.*, 2000b) as well as in reducing listener perception of magnitude of modulation of voice during connected speech (Hertegård *et al.*, 2000) for some individuals with laryngeal vocal tremor. However, it is unknown whether or not botulinum toxin injections actually reduce the extent of F0 and intensity modulation, or reduce the perception of modulation based on acoustical changes as seen in the current study.

Further studies are needed in humans to determine if volitional production of a breathy voice can reduce the perception of magnitude of modulation of voice as it did with

TABLE III. Results of study 2 binary logistic stepwise regression for each contrast (F0, degree of vocal fold adduction, vocal tract shape) with potential predictor variables representing (1) the ratios of the mean amplitudes in each of the six octave bands for each of the stimulus pairs and (2) the ratios of the coefficients of variation of amplitude in each of the six octave bands (center frequencies: 125, 250, 500, 1000, 2000, and 4000 Hz) for each of the stimulus pairs for the glottal noise and no glottal noise conditions.

			125 Hz	250 Hz	500 Hz	1000 Hz	2000 Hz	4000 Hz	Cox and Snell R-square
Glottal Noise	Mean Amplitude	F0	1	—	—	—	—	—	0.004, χ^2 (1, N = 2688) = 10.953, $p < 0.005$
		Adduction	1	—	3	4	2	—	0.215, χ^2 (4, N = 2688) = 649.821, $p < 0.001$
		Shape	—	—	—	1	—	—	0.033, χ^2 (1, N = 2688) = 90.498, $p < 0.001$
	Coefficient of Variation	F0	2	—	1	—	—	—	0.008, χ^2 (2, N = 2688) = 20.660, $p < 0.05$
		Adduction	—	3	1	4	2	5	0.151, χ^2 (5, N = 2688) = 441.596, $p < 0.001$
		Shape	3	—	2	1	—	4	0.024, χ^2 (4, N = 2688) = 64.381, $p < 0.001$
No Glottal Noise	Mean Amplitude	F0	2	—	—	—	—	1	0.025, χ^2 (2, N = 2688) = 68.598, $p < 0.001$
		Adduction	1	—	2	—	—	—	0.043, χ^2 (2, N = 2688) = 117.215, $p < 0.001$
		Shape	—	—	—	1	—	—	0.016, χ^2 (1, N = 2688) = 44.343, $p < 0.001$
	Coefficient of Variation	F0	1	—	—	2	—	—	0.012, χ^2 (2, N = 2688) = 32.089, $p < 0.001$
		Adduction	2	3	1	—	—	4	0.022, χ^2 (4, N = 2688) = 59.399, $p < 0.001$
		Shape	—	—	1	—	—	2	0.005, χ^2 (2, N = 2688) = 14.429, $p < 0.005$

computer-simulated vocal tremor. If it does, the acoustical bases for the perceptual results should be determined and compared to the findings of the current study. Because individuals with essential tremor present with tremor affecting multiple components of the speech mechanism, a human model of vocal tremor (i.e., a singer producing vibrato) might allow for better isolation of the source of modulation to the larynx to investigate these questions.

IV. CONCLUSIONS

The use of computer-simulated voice samples involving primary modulation of the F0 revealed that adjustments to the voice source (i.e., vocal fold length, degree of vocal fold adduction) and vocal tract filter (i.e., vocal tract configuration for vowels) affect the perception of the magnitude of voice modulation. Specifically, listeners perceived a higher magnitude of voice modulation when simulated samples had a higher mean F0 corresponding to an increased vocal fold length, greater degree of vocal fold adduction corresponding to a pressed voice, and a vocal tract shape with clustering of resonant peaks in higher frequency regions (i.e., /i/ vs /a/). However, the effect of F0 was significant only when glottal noise was not present in the acoustic signal, which is not typical in humans. Based on acoustical analyses, listeners may use combinations of low and high frequency information to make judgments about the magnitude of modulation of voice. However, the acoustical measures used in this study did not fully account for listeners' judgments. Further studies are needed to investigate the acoustical information and the strategies that listeners use to judge the magnitude of modulation of complex sounds and to determine if the results from this simulation study hold true in humans.

ACKNOWLEDGMENTS

This research was funded by National Institutes of Health Grant No. F31 DC012697 and R01-DC011275. We would like to thank Andrew Lotto and Mark Borgstrom for their assistance with the statistical analyses for this study.

- Adler, C. H., Bansberg, S. F., Hentz, J. G., Ramig, L. O., Buder, E. H., Witt, K., Edwards, B. W., Krein-Jones, K., and Caviness, J. N. (2004). "Botulinum toxin type A for treating voice tremor," *Arch. Neurol.* **61**(9), 1416–1420.
- American Speech-Language-Hearing Association. (1997). *Guidelines for Audiological Screening*, Rockville, MD.
- ANSI (1993). S1.1-1986 (ASA 65-1986), "Specifications for Octave-Band and Fractional-Octave-Band Analog and % Digital Filters" (American National Standards Institute, New York).
- Barkmeier-Kraemer, J., Lato, A., and Wiley, K. (2011). "Development of a speech treatment program for a client with essential vocal tremor," *Sem. Speech Lang.* **32**(1), 43–57.
- Brown, J. R., and Simonson, J. (1963). "Organic voice tremor: A tremor of phonation," *Neurology* **13**(6), 520–520.
- Bunton, K., and Story, B. H. (2010). "Identification of synthetic vowels based on a time-varying model of the vocal tract area function," *J. Acoust. Soc. Am.* **127**(4), EL146–EL152.
- Burgen, A., Dickens, F., and Zatman, L. (1949). "The action of botulinum toxin on the neuro-muscular junction," *J. Physiol.* **109**, 10–24.
- Cohen, J. (1969). *Statistical Power Analysis for the Behavioral Sciences* (Academic, New York).
- Deuschl, G., Bain, P., and Brin, M. (1998). "Consensus statement of the movement disorder society on tremor," *Move. Disord.* **13**(S3), 2–23.
- Dromey, C., Warrick, P., and Irish, J. (2002). "The influence of pitch and loudness changes on the acoustics of vocal tremor," *J. Speech Lang. Hear. Res.* **45**(5), 879–890.
- Fant, G. (1960). *The Acoustic Theory of Speech Production* (Moulton, The Hague), pp. 15–21.
- Farinella, K., Hixon, T., Hoit, J., Story, B., and Jones, P. (2006). "Listener perception of respiratory-induced voice tremor," *Am. J. Speech-Lang. Pathol.* **15**, 72–84.
- Faul, F., Erdfelder, E., Lang, A.-G., and Buchner, A. (2007). "G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences," *Behav. Res. Meth.* **39**(2), 175–191.
- Gamboa, J., Jiménez-Jiménez, F. J., Nieto, A., Cobeta, I., Vegas, A., Ortí-Pareja, M., Gasalla, T., Molina, J. A., and García-Albea, E. (1998). "Acoustic voice analysis in patients with essential tremor," *J. Voice* **12**(4), 444–452.
- Hertegård, S., Granqvist, S., and Lindstad, P. (2000). "Botulinum toxin injections for essential voice tremor," *Ann. Otol. Rhinol. Laryngol.* **109**(2), 204–209.
- Hillenbrand, J. M., and Gayvert, R. T. (2005). "Open source software for experiment design and control," *J. Speech Lang. Hear. Res.* **48**(1), 45–60.
- Hollien, H., Mendes-Schwartz, A. P., and Nielsen, K. (2000). "Perceptual confusions of high-pitched sung vowels," *J. Voice* **14**(2), 287–298.
- Horii, Y., and Hata, K. (1988). "A note on phase relationships between frequency and amplitude modulations in vocal vibrato," *Folia Phoniatr. Logop.* **40**(6), 303–311.
- Koller, W. C., Busenbark, K., and Miner, K. (1994). "The relationship of essential tremor to other movement disorders: Report on 678 patients," *Ann. Neurol.* **35**(6), 717–723.

- Koller, W. C., Glatt, S., Biary, N., and Rubino, F. A. (1987). "Essential tremor variants: Effect of treatment," *Clin. Neuropharm.* **10**(4), 342–350.
- Koller, W., Graner, D., and Mlcoch, A. (1985). "Essential voice tremor: Treatment with propranolol," *Neurology* **35**(1), 106–106.
- Kreiman, J., and Gerratt, B. R. (2012). "Perceptual interaction of the harmonic source and noise in voice," *J. Acoust. Soc. Am.* **131**(1), 492–500.
- Lederle, A., Barkmeier-Kraemer, J., and Finnegan, E. (2012). "Perception of vocal tremor during sustained phonation compared with sentence context," *J. Voice* **26**(5), 668.e1–668.e9.
- Lester, R. A. (2014). "Modulation of voice related to tremor and vibrato," Ph.D. Dissertation, University of Arizona, Tucson, AZ.
- Lester, R. A., Barkmeier-Kraemer, J., and Story, B. H. (2013). "Physiologic and acoustic patterns of essential vocal tremor," *J. Voice* **27**(4), 422–432.
- Lester, R. A., and Story, B. H. (2013). "Modulation of voice related to simulated vocal fold length change with cricothyroid and thyroarytenoid muscle activation," in *10th International Conference Advances in Quantitative Laryngology, Voice and Speech Research*, pp. 63, 64.
- Liljencrants, J. (1985). *Speech Synthesis with a Reflection-type Line Analog*, Royal Institute of Technology, Department of Speech Communication and Music Acoustics, Stockholm, Sweden.
- Lou, J.-S., and Jankovic, J. (1991). "Essential tremor: Clinical correlates in 350 patients," *Neurology* **41**(2 Part 1), 234–234.
- Louis, E. D. (2005). "Essential tremor," *Lancet Neurol.* **4**(2), 100–110.
- Ludlow, C. L., Bassich, C. J., Connor, N. P., and Coulter, D. C. (1986). "Phonatory characteristics of vocal fold tremor," *J. Phonetics* **14**, 509–515.
- Ludlow, C. L., Naunton, R. F., Fujita, M., and Sedory, S. E. (1990). "Spasmodic dysphonia: Botulinum toxin injection after recurrent nerve surgery," *Otolaryng. Head Neck Surg.* **102**(2), 122–131.
- Massey, E. W., and Paulson, G. W. (1985). "Essential vocal tremor: Clinical characteristics and response to therapy," *South. Med. J.* **78**(3), 316–317.
- Peterson, G. E., and Barney, H. L. (1952). "Control methods used in a study of the vowels," *J. Acoust. Soc. Am.* **24**(2), 175–184.
- Ramig, L. A., and Shipp, T. (1987). "Comparative measures of vocal tremor and vocal vibrato," *J. Voice* **1**(2), 162–167.
- Samlan, R. A., and Story, B. H. (2011). "Relation of structural and vibratory kinematics of the vocal folds to two acoustic measures of breathy voice based on computational modeling," *J. Speech Lang. Hear. Res.* **54**(5), 1267–1283.
- Story, B. H. (1995). "Physiologically-based speech simulation using an enhanced wave-reflection model of the vocal tract," Ph.D. Dissertation, University of Iowa, Iowa City, IA.
- Story, B. H. (2005). "Synergistic modes of vocal tract articulation for American English vowels," *J. Acoust. Soc. Am.* **118**(6), 3834–3859.
- Story, B. H. (2013). "Phrase-level speech simulation with an airway modulation model of speech production," *Comput. Speech Lang.* **27**(4), 989–1010.
- Story, B. H., Titze, I. R., and Hoffman, E. A. (1996). "Vocal tract area functions from magnetic resonance imaging," *J. Acoust. Soc. Am.* **100**(1), 537–554.
- Sundberg, J. (1974). "Articulatory interpretation of the 'singing formant,'" *J. Acoust. Soc. Am.* **55**, 838–844.
- Titze, I. R. (1984). "Parameterization of the glottal area, glottal flow, and vocal fold contact area," *J. Acoust. Soc. Am.* **75**, 570.
- Titze, I. R. (1994). *Principles of Voice Production* (Prentice-Hall, Englewood Cliffs, NJ), pp. 120, 248.
- Titze, I. R. (2006). *The Myoelastic Aerodynamic Theory of Phonation* (National Center for Voice and Speech, Iowa City, IA), pp. 197–214.
- Warrick, P., Dromey, C., Irish, J., and Durkin, L. (2000b). "The treatment of essential voice tremor with botulinum toxin A: A longitudinal case report," *J. Voice* **14**(3), 410–421.
- Warrick, P., Dromey, C., Irish, J. C., Durkin, L., Pakiam, A., and Lang, A. (2000a). "Botulinum toxin for essential tremor of the voice with multiple anatomical sites of tremor: A crossover design study of unilateral versus bilateral injection," *The Laryngoscope* **110**(8), 1366–1374.
- Wegel, R. L., and Lane, C. E. (1924). "The auditory masking of one pure tone by another and its probable relation to the dynamics of the inner ear," *Phys. Rev.* **23**(2), 266–285.
- Weiss, R., Brown, W., and Moris, J. (2001). "Singer's formant in sopranos: Fact or fiction?," *J. Voice* **15**(4), 457–468.