# An Acoustic-Perceptual Study of Vocal Tremor

*Supraja Anand, †Rahul Shrivastav, *Judith M. Wingate, and ‡Neil N. Chheda, *‡Gainesville, Florida and †East Lansing, Michigan

**Summary: Background.** Essential tremor of the voice (ETV) is an involuntary intention tremor of the vocal folds that causes fluctuations in fundamental frequency ($f_0$) and/or intensity leading to an unsteady voice. There is limited data on how different acoustic variables affect perception of severity of tremor.

**Aim.** The purpose of the study was to determine if systematic changes in $f_0$, rate or modulation frequency ($f_{f0m}$), extent or depth of modulation ($d_{f0m}$), and signal-to-noise ratio (SNR) affect perception of severity of tremor.

**Method.** Vowel phonations of four speakers (two male and two female) with a clinical diagnosis of ETV were selected from the Kay Elemetrics Disordered Voice Database (Lincoln Park, NJ). A high fidelity speech vocoder (STRAIGHT; Kawahara, 1997) was used to synthesize the $f_0$ contour for each of these voices, which were varied in mean $f_0$, $f_{f0m}$, and $d_{f0m}$. The $f_0$ contour was modified 30 Hz above and below the mean $f_0$ for each speaker. $f_{f0m}$ ranged from 3 to 12 Hz in steps of 3 Hz. $d_{f0m}$ ranged from 2 to 32 Hz in steps of 6 Hz. Six (three experts and three naïve) listeners rated the "severity" of tremor on a seven-point rating scale.

**Results.** Significant main effects and interactions were found between the study variables. Perceived severity of tremor increased with $f_{f0m}$ and $d_{f0m}$. There was no systematic effect of SNR on perceived tremor severity.

**Conclusion.** The perception of severity for steady-state tremor results from a complex interaction of multiple acoustic cues with $d_{f0m}$ acting as the primary acoustic cue.

**Key Words:** Essential tremor of the voice (ETV)–Perception–Speech synthesis–Modulation frequency ($f_{f0m}$)–Modulation depth ($d_{f0m}$).

## INTRODUCTION

Essential tremor is one of the most common movement disorders. Tremor has been defined as an "involuntary, approximately rhythmic and roughly sinusoidal movement."[1] It could be caused by a multitude of factors such as neurological disorders (eg, Parkinson's disease, multiple sclerosis), withdrawal from alcohol or addictive drugs, hereditary causes, or be of idiopathic origin. The upper extremities are usually most affected, but tremor can also occur in other body parts including the head, trunk, lower extremities, and the larynx.[2–4] Essential tremor of the voice (ETV) is a "progressive intention tremor" of the vocal folds, which causes fluctuations in fundamental frequency ($f_0$) and/or intensity. The result of this disturbance is an unsteady voice.[5–7] Tremor of other sites besides the larynx including the palate, back of the tongue, and pharynx may co-occur with vocal fold tremor and may have an impact on voice.[4,8]

Essential tremor of the voice has been treated pharmacologically by a number of agents such as antiepileptics (primidone), beta-blockers (propranolol), benzodiazepines (clonazepam, diazepam), and carbonic anhydrase inhibitors (methazolamide). Nevertheless, results from the above studies have concluded that patients with vocal tremor receive only a modest benefit with response rates ranging from 25% to 40%.[9–14] Ever since botulinium toxin (BOTOX) was introduced as a therapeutic agent, it has become the primary treatment for ETV. BOTOX reduces the perceived effort to speak, and the rate and extent of tremor.[15–17] BOTOX provides symptomatic relief and can be useful only when the tremor is primarily linked to the vocal folds because it is commonly injected into the thyroarytenoid muscle. Warrick et al[5] showed that when tremor involved several anatomical sites, injections of BOTOX did not lead to uniform improvements on acoustic measures of voice. Success rates for vocal tremor treated with BOTOX have been only 50–65% on the basis of perceptual and acoustic measures.[5,16] In addition to this relatively low success rate of BOTOX, side effects of breathiness and dysphagia have also been reported in the literature.[5,17]

A large number of experiments have been carried out to characterize the tremulous voice quantitatively. These studies have examined frequency modulation characteristics (rate and extent) and short-term $f_0$ variations such as jitter through either time domain representations[18–22] or frequency domain representations.[6] Tremor rate is defined as the rate of modulation occurring about the mean $f_0$ (hereafter, described as $f_{f0m}$ or as modulation frequency). Results from the above-mentioned studies concluded that rate of $f_0$ modulation in patients with ETV varied between 2 and 12 Hz with high interspeaker variability. Tremor extent can be described by how large or small the modulations are relative to the mean $f_0$. The extent of $f_0$ modulation or modulation depth ($d_{f0m}$) ranged from 0.5 to 2 semitones (STs).[21,23]

Studies measuring $f_{f0m}$ and $d_{f0m}$ provide only a partial insight into vocal tremor, because these do not describe the perceptual significance of these acoustic features completely. A better understanding of acoustic-perceptual relationship of ETV is critical because majority of the clinical decisions are heavily influenced by their perceptual effects. For example, dosage levels of BOTOX injections are chosen based on perceived reduction of tremor severity. Yet, only few studies have systematically investigated the relationship between

Accepted for publication February 27, 2012.
Presented at the 38th Annual Symposium of the Voice Foundation, 2009 (poster format).
From the *Department of Speech, Language, and Hearing Sciences, University of Florida, Gainesville, Florida; †Department of Communicative Sciences and Disorders, Michigan State University, East Lansing, Michigan; and the ‡Department of Otolaryngology, University of Florida, Gainesville, Florida.
Address correspondence and reprint requests to Supraja Anand, Department of Speech, Language, and Hearing Sciences, University of Florida, Gainesville, FL 32611. E-mail: supraja.anand@ufl.edu
Journal of Voice, Vol. 26, No. 6, pp. 811.e1-811.e7
0892-1997/$36.00

doi:10.1016/j.jvoice.2012.02.007

acoustic parameters and the perception of vocal tremor in pathologic voices.

One study that examined the effects of regularity of $f_{f0m}$ and $d_{f0m}$ on the perception of vocal tremor was done by Kreiman et al.[24] They used a custom-designed speech synthesizer to model voices with vocal tremor. This voice synthesizer was used to create sinusoidal and irregular modulations of vocal tremor. Vowel phonations were synthesized using different rate and extent parameters. Modulation frequency ($f_{f0m}$) was varied between 4 and 20 Hz. The $d_{f0m}$, computed as the average deviation in $f_0$ around the mean $f_0$, was varied from 0.6 to 10.3 Hz. Listeners compared similarity between original stimuli and the synthesized versions on a 100-point visual analog scale. Their findings revealed that listeners judged the synthetic stimuli to be very similar to their original targets (sinusoidal model: $25.6 \pm 26.3$; irregular model: $24.1 \pm 25.9$). Listeners were insensitive to the precise details of the $f_0$ contour when severity of vocal tremor was mild. Significant interactive effects were found between $f_{f0m}$, $d_{f0m}$, and regularity on the perception of tremor severity. Differences in tremor rate were easiest to hear when the tremor was sinusoidal and of small amplitude. Listeners had more difficulty focusing on the tremor rate alone when the extent of tremor increased because the complexity of tremor pattern increased or when the tremor model was irregular. Results from this study suggested that the listeners were not sensitive to relatively small differences in amount of $d_{f0m}$. Hence, a systematic manipulation of $f_{f0m}$ and $d_{f0m}$ is needed to explain the perceptual findings.

Past experiments that explored the relationship between acoustic and perceptual measures of voice have often assumed a direct and linear relationship between the two. However, the relationship between a physical stimulus and its perceptual attribute is often nonlinear.[25–27] Just as the perception of pitch, intensity, and quality of the voice are nonlinear functions of specific changes to the vocal acoustic signal,[25–27] perceived change in tremor severity may also be cued by multiple acoustic changes (eg, the rate, extent, and regularity of modulation), interacting in a complex nonlinear manner.

To identify such relationships, the present study systematically manipulated $f_0$, $f_{f0m}$, $d_{f0m}$, and signal-to-noise ratio (SNR) and evaluated their effects on the perception of tremor severity. In human listeners, amplitude modulated carrier signals can elicit a range of percepts including wobble ($f_{f0m} < 4$ Hz), vibrato ($f_{f0m}$ between 4 and 7 Hz),[21,23] tremor ($f_{f0m}$ between 2 and 12 Hz),[6,24] and roughness ($f_{f0m}$ between 20 and 40 Hz).[28–30] These studies have shown $d_{f0m}$ to affect the perception of characteristics such as roughness.[29–31] Such findings from the psychoacoustic literature form the basis for the following hypotheses:

1. The perceived severity of essential voice tremor would increase with an increase in $d_{f0m}$,
2. The $f_{f0m}$ that results in the greatest perceived severity of essential voice tremor will vary with the mean $f_0$ of that voice, and
3. There may be an interaction between severity of dysphonia (measured using SNR) and the perception of tremor severity. This interaction may result in greater perceived tremor severity for voices with greater dysphonia (low SNR).

## METHOD

### Speakers

Phonation samples from two male and two female speakers were selected from a large database of disordered voices (Kay Elemetrics Disordered Voice Database, Lincoln Park, NJ).[32] The four speakers were diagnosed as having "ETV." Voice analysis software *TF32* (Milenkovic, Madison, WI) was used to obtain the SNR values.[33] The SNR for the four speakers was used as a gross metric of dysphonia severity, and voices with high SNR were assumed to have relatively good voice quality, whereas those with low SNR were assumed to indicate dysphonia. Details about these speakers are provided in Table 1. The primary objective in this study was to test speakers who represented a wide range of $f_0$ and voice quality. This was achieved through careful selection of the four speakers with varying $f_0$ and SNRs. As the goal was to evaluate how tremor perception changes based on specific signal characteristics, the exact number of speakers was not deemed to be as critical as the range of characteristics under study.

### Stimuli

A 1-second sample from temporal midpoint of the vowel phonation /a/ was extracted. Inclusion of longer segments increases the risk of voice quality variation within each stimulus and can complicate interpretation of results. Furthermore, there is little reason to believe that 1-second samples limit the perception of tremor in anyway as prior study on perception of vocal tremor had used 1-second samples for perceptual analysis.[24] An algorithm was designed in *MATLAB* (version 7.0; MathWorks, Natick, MA) to systematically manipulate the mean $f_0$, $f_{f0m}$, and $d_{f0m}$ of tremor, through the use of a high fidelity speech vocoder (STRAIGHT; Kawahara, 1997).[34] The $f_{f0m}$ was varied from 3 to 12 Hz in steps of 3 Hz. The $d_{f0m}$ was varied from 2 to 32 Hz in steps of 6 Hz. These ranges were chosen to create a continuum of stimuli, which ranged from low to high $f_{f0m}$ and $d_{f0m}$, in accordance with the values reported in the literature.[4,7,18,20,24] Each stimulus was further modified to scale the mean $f_0$ 30 Hz above and below the natural $f_0$ for that speaker. Thus, a total of 288 stimuli (4 speakers $\times$ 3 mean $f_0 \times 4$ $f_{f0m} \times 6$ $d_{f0m}$) were created. Figure 1 depicts an example of variations

**TABLE 1.**
**Subject Characteristics**

| Subject ID | Age (y) | Gender | Fundamental Frequency (Hz) | Signal-to-Noise Ratio (dB) |
|---|---|---|---|---|
| S01 | 76 | M | 157 | 20.3 |
| S02 | 69 | M | 128 | 9.7 |
| S03 | 78 | F | 212 | 20.5 |
| S04 | 73 | F | 206 | 13.3 |

in $f_{f0m}$ and $d_{f0m}$ for a stimulus with mean $f_0$ of 128 Hz. All the stimuli were resampled to 24,414 Hz to match the hardware requirements of the equipment used for the perceptual test.

As perception of pitch is logarithmically related to the $f_0$, the mean $f_0$ of all stimuli were converted to a ST scale using the formula given by Baken and Orlikoff.[35] These ST values were used for all subsequent statistical analyses.

$$\mathrm{ST} = 39.86 \times \log10\left(\frac{f_0}{f_{0\mathrm{ref}}}\right) \quad (1)$$

For these calculations, the lowest $f_0$ among the four speakers (98 Hz) was selected as the reference frequency ($f_{0\mathrm{ref}}$).

### Listeners

Two speech-language pathologists and one ear, nose, and throat specialist (two females and one male, respectively) with a minimum of 20 years of experience in the diagnosis and rehabilitation of patients with voice disorders served as expert listeners. Three female undergraduate students from Department of Speech, Language, and Hearing Sciences served as naïve listeners in this study. The mean age of expert listeners was 47 years and the mean age of naïve listeners was 23 years. All the listeners were native speakers of American English and passed a hearing screening (air-conduction pure tone threshold below 20 dB hearing level or HL at 250, 500, 1000, 2000, and 4000 Hz) before the listening task.[36]

### Instrumentation

All the data acquisition procedures were controlled using the software, *Sykofizx* and the TDT System III hardware (Tucker-Davis Technologies, Inc., Alachua, FL). Stimuli were presented to the listeners in a single-walled sound booth using ER2 insert earphones (Etymotic Research, Inc., Elk Grove Village, IL). These earphones were chosen for their flat frequency response at the tympanic membrane. Stimuli were presented at 85 dB sound pressure level or SPL in the right ear to avoid potential effects of binaural interaction.

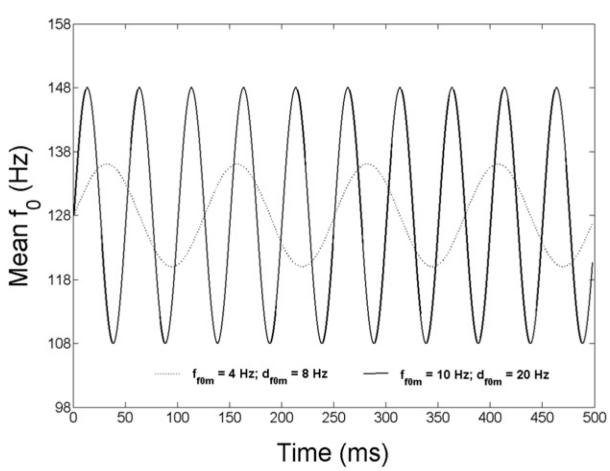


**FIGURE 1.** A graphic representation of $f_{f0m}$ and $d_{f0m}$ variations for a stimulus with mean $f_0$ of 128 Hz.

### Procedure

Listeners rated the severity of tremor for each stimulus on a seven-point rating scale (1 = no tremor, 7 = severe tremor) using a custom-designed interface. Shrivastav et al[37] showed that interlistener variability in rating scale data was minimized when multiple ratings of a stimulus were averaged and standardized. Hence, multiple ratings were obtained from each listener and averaged to compute a single score for each stimulus. Each stimulus was presented a total of five times in random order for a total of 1440 test stimuli. Given the large number of stimuli, listeners were required to take several short breaks during the listening task to avoid fatigue and maintain their attention. Testing took approximately 2 hours for each listener. Testing was completed in a single test session.

### Statistical analyses

All the statistical procedures were carried out using *SPSS* version 18.0 (SPSS Inc., Chicago, IL). An independent sample *t* test was performed to compare the perceived tremor severity scores between the two listener groups (naïve and experts). To determine the effects of $f_0$, $f_{f0m}$, $d_{f0m}$, and SNR on perception of tremor severity, a four-way analysis of variance (ANOVA) with post hoc comparisons using Boniferroni's correction was used. Mean $f_0$, $f_{f0m}$, $d_{f0m}$, and SNR were the independent variables in the present study. The average tremor severity judgment obtained from all listeners served as the dependent variable. In addition, a multiple linear regression analysis was performed to develop a model with the three predictor variables of tremor severity. Different regression functions were derived to fit the perceptual data and explain the relationship between the independent variables and perceived tremor severity.

## RESULTS

### Reliability analysis

Interjudge reliability was defined as the degree of consistency between average ratings of tremor severity between listeners for all the stimuli. Intrajudge reliability was defined as the degree of consistency within listeners between the five trials of tremor severity rating for all the stimuli. Results from intraclass correlation coefficient revealed that reliability between listeners was 0.7 and reliability within listeners was 0.9. These high correlations revealed that listeners made consistent judgments of tremor severity.

### Group differences

Table 2 depicts the mean perceived tremor severity scores for both listener groups (naïve and experts). An independent sample *t* test revealed that there was a significant difference ($t$ (2878) = 2.332, $P = 0.02$) between the two listener groups in mean tremor severity ratings. Naïve listeners rated the voices to have greater tremor severity than expert listeners. However, the magnitude of difference was found to be very small with a mean difference of 0.12. This difference represents only 2.27% of the total range of responses obtained from listeners. As the purpose of this study was investigation of trends

**TABLE 2.**
**Mean and SD Scores for Naïve and Expert Listeners**

| Listener Type | N | Mean Severity | SD |
|---|---|---|---|
| Naive | 1440 | 4.4884 | 1.53275 |
| Expert | 1440 | 4.3648 | 1.30196 |

*Abbreviation:* SD, standard deviation.

(ie, how one variable is affected by another), we combined data across the two groups for ANOVA testing.

## Main effects

Significant main effect was found for mean $f_0$ ($F(11, 2592) = 92.77$, $P < 0.001$) as illustrated in Figure 2. Voices with low $f_0$ were perceived to have greater tremor severity compared with those with higher $f_0$. Significant main effects were also found for $f_{f0m}$ ($F(3, 2592) = 74.18$, $P < 0.001$) and $d_{f0m}$ ($F(5, 2592) = 694.26$, $P < 0.001$) as illustrated in Figures 3 and 4, respectively. On average, higher modulation frequencies were judged to have greater tremor severity when compared to lower modulation frequencies (Figure 3). Figure 4 shows that the perceived tremor severity increases with an increase in $d_{f0m}$, but this relationship appears to be nonlinear with the greatest change in tremor severity occurring for $d_{f0m}$ below 20 Hz. There was also a significant main effect of SNR ($F(57, 518) = 272.01$, $P < 0.001$) as illustrated in Figure 5. SNR values for the synthetic stimuli ranged from 7.0 to 21.0. However, this effect was not systematic.

## Interaction effects

**Interaction 1: Mean $f_0$ vs $f_{f0m}$.** A significant interaction was observed between mean $f_0$ and $f_{f0m}$ ($F(33, 2592) = 1.88$, $P < 0.001$) as illustrated in the first row of plots in Figure 6. Perceived severity of tremor decreased with an increase in mean $f_0$ for stimuli that have low fundamental frequencies (approximately 180 Hz and below). However, it was only a small decrease (±0.5) in mean tremor severity ratings. At all mean $f_0$, $f_{f0m}$ of 12 Hz was perceived to have greater tremor severity.

**Interaction 2: Mean $f_0$ vs $d_{f0m}$.** A significant interaction was also found between mean $f_0$ and $d_{f0m}$ ($F(55, 2592) = 1.82$, $P < 0.001$) as illustrated in the second row of plots in Figure 6. Similar to the $f_0$-$f_{f0m}$ interaction, perceived severity of tremor decreased with an increase in mean $f_0$ for stimuli that have low fundamental frequencies (approximately 180 Hz and below). For all mean $f_0$, $d_{f0m}$ of 2 Hz was perceived to have the least tremor severity and $d_{f0m}$ of 32 Hz was perceived to have the greatest tremor severity.

**Interaction 3: $f_{f0m}$ vs $d_{f0m}$.** A final significant interaction was found between $f_{f0m}$ and $d_{f0m}$ ($F(15, 2592) = 6.60$,

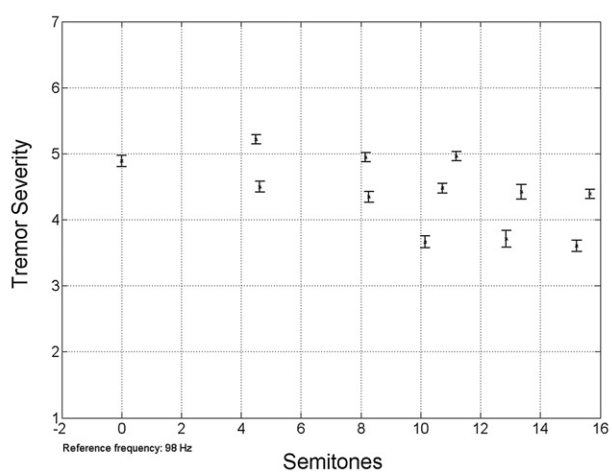


**FIGURE 2.** Perceived tremor severity plotted against the mean $f_0$ in semitone scale with the reference frequency of 98 Hz. Symbols indicate the perceived tremor severity averaged across listeners and bars indicate standard error of the mean (SEM).

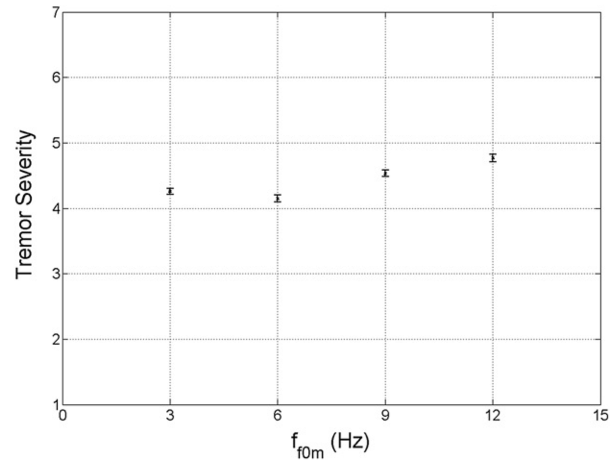


**FIGURE 3.** Effect of $f_{f0m}$ on perceived tremor severity. Symbols indicate the perceived tremor severity averaged across listeners and bars indicate standard error of the mean (SEM).

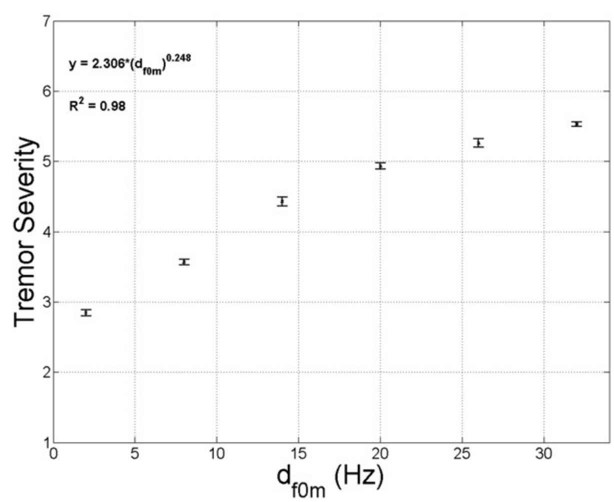


**FIGURE 4.** Effect of $d_{f0m}$ on perceived tremor severity. Symbols indicate the perceived tremor severity averaged across listeners and bars indicate standard error of the mean (SEM).

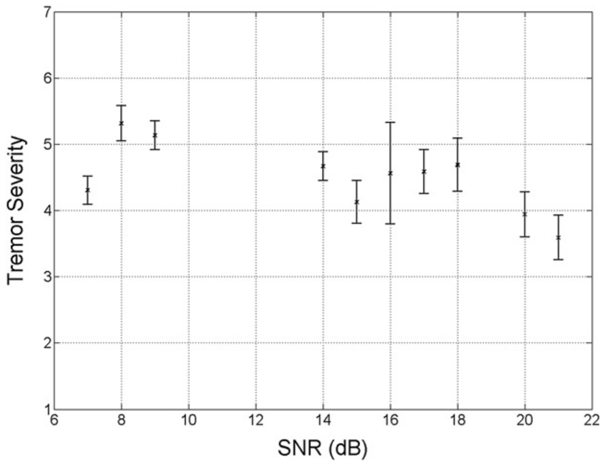


**FIGURE 5.** Effect of SNR on perceived tremor severity. Symbols indicate the perceived tremor severity averaged across listeners and bars indicate standard error of the mean (SEM). The SEM for an SNR of 16 dB was large because of a small number of samples (N).

$P < 0.001$). The mean tremor severity scores increased with $d_{f0m}$ for all modulation frequencies. Although all $f_{f0m}$ elicited the same mean tremor severity at low $d_{f0m}$, stimuli with $d_{f0m}$s above 8 Hz show greater change in tremor severity for higher modulation frequencies. This is illustrated in Figure 7. In other words, the effects of $f_{f0m}$ became more evident only for $d_{f0m}$ greater than 8 Hz. Below this $d_{f0m,}$ all $f_{f0m}$ appear to affect tremor severity equally.

## Regression analysis

Multiple linear regression analysis was carried out to relate all independent variables (mean $f_0$, $f_{f0m}$, $d_{f0m}$) to the perceived tremor severity. The equation based on the unstandardized coefficient of the regression for the entire data set is:

$$y = 3.652 - 0.007 f_0 + 0.063 f_{f0m} + 0.091 d_{f0m} \tag{2}$$

The overall variance accounted for perceived tremor severity amounted to 49.3%.

In addition, there was a clear nonlinear relationship between $d_{f0m}$ and perceived tremor severity. Therefore, a nonlinear regression analysis was conducted to determine the best fit function to explain how $d_{f0m}$ may be related to perceived tremor severity. The goodness of fit as shown in Figure 4 was obtained using a power function:

$$y = 2.306 * \left(d_{f0m}\right)^{0.248} \tag{3}$$

This equation shows the effect of $d_{f0m}$ on perceived tremor severity when averaged across all mean $f_0$ and $f_{f0m.}$ The proportion of variance obtained using a power function ($R^2 = 0.98$) improved compared with the linear fit.

To explain the interaction between $f_{f0m}$ and $d_{f0m}$ as shown in Figure 7, a power function was used to fit the perceptual data

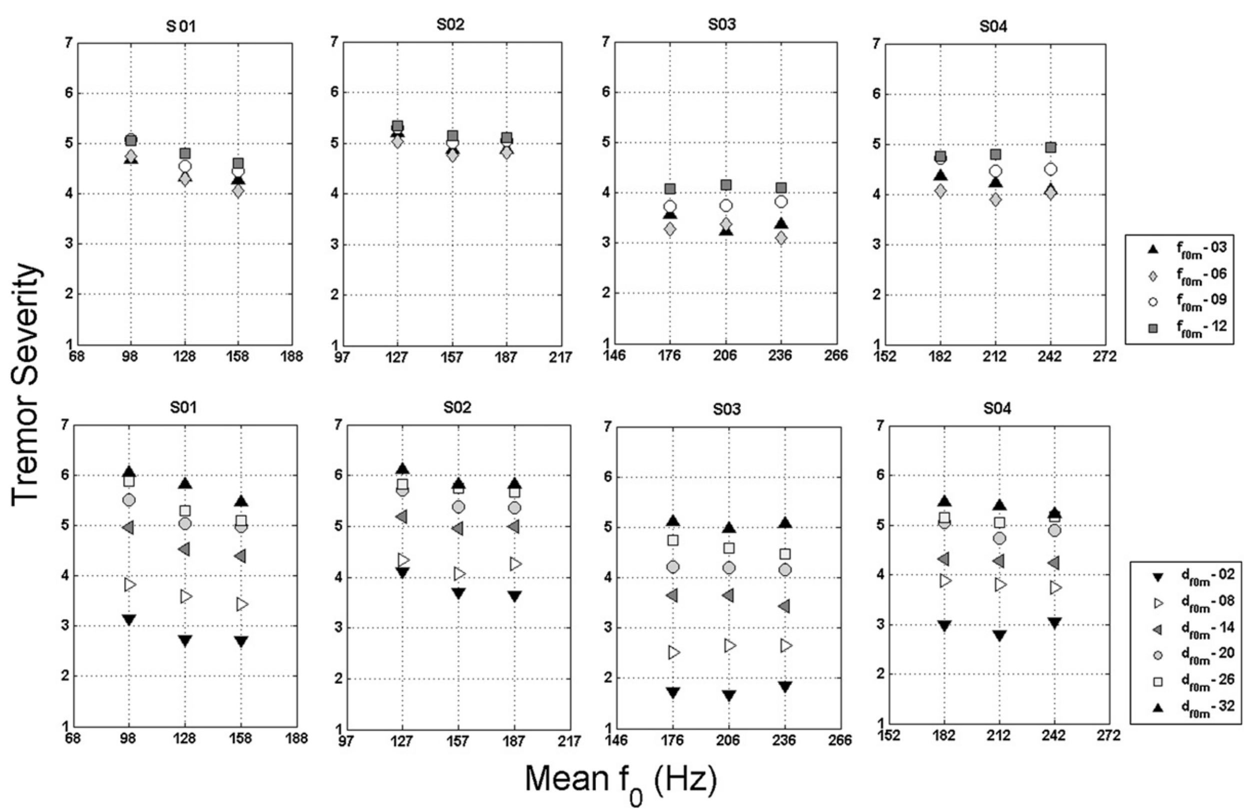


**FIGURE 6.** The tremor severity judgments are shown as a function of mean $f_0$ (in Hz) for each of the four speakers. The top row indicates interaction between mean $f_0$ and $f_{f0m}$ and the bottom row indicates interaction between mean $f_0$ and $d_{f0m}$.

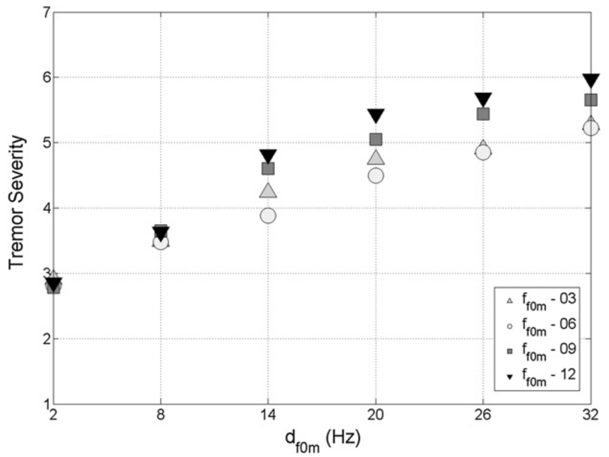


**FIGURE 7.** The tremor severity judgments are shown as a function of $d_{f0m}$ (in Hz) for the different $f_{f0m}$ values.

with different powers representing the various $f_{f0m}$ ($\beta = 0.239$: 3 Hz; $\beta = 0.247$: 6 Hz; $\beta = 0.288$: 9 Hz; $\beta = 0.303$: 12 Hz).

$$y = 2.119 * \left(d_{f0m}\right)^{P:\,0.239-0.303} \quad (4)$$

## DISCUSSION

Essential tremor of the voice can be considered as the laryngeal manifestation of essential tremor. ETV refers to the periodic contraction of antagonistic adductor-abductor and/or superior-inferior laryngeal muscles in an alternating or synchronous fashion. Symptoms are progressive and potentially disabling. Although a number of medications have been used to treat tremor in the past, BOTOX has been the primary choice of treatment.[5,15–17] Extensive research has been carried out on the acoustic characteristics of vocal tremor. However, clinical judgments of voice disorders by patients and therapists rely heavily on the perception of tremor in voice. Till date, only Kreiman et al[24] have attempted to relate acoustics to perception; however, the range of modulation parameters tested in this study failed to produce significant perceptual changes.

This study sought to explore how variations in $f_0$, $f_{f0m}$, $d_{f0m}$, and SNR may affect the perceived tremor severity through systematic manipulation of these independent variables. The findings from the present study confirmed the first hypothesis and suggested that perceived severity of tremor increases with an increase in $d_{f0m}$. Also, interaction effects ($f_{f0m}$ vs $d_{f0m}$) become evident after 8 Hz. These results are based on categorical scaling of tremor severity. In contrast, Kreiman et al[24] performed a discrimination experiment, which suggested that listeners are not sensitive to small differences in tremor amplitude. So, there may be a threshold that must be reached in order for listeners to discern significant differences in $d_{f0m}$. Hence, the results of the present study cannot be directly compared with Kreiman et al[24] owing to the difference in goals and experimental procedures.

Perceived tremor severity decreased with an increase in mean $f_0$ at low frequencies only. From Figure 6, it can be observed that perceived mean severity ratings decreased only for mean $f_0$ values in S01 and S02, where $f_0$ ranged from 98 to 158 Hz and 127 to 187 Hz, respectively. However, the magnitude of decrease in mean tremor severity ratings was small (±0.5 units). The interactive effects between mean $f_0$ and $f_{f0m}$ are evident only for higher mean $f_0$ values (S03 and S04), thus supporting hypothesis #2.

The results of the present study are in agreement with psychoacoustic studies that have shown $d_{f0m}$ as one of the primary parameters influencing the voice quality "roughness." Perceptual mechanisms underlying vocal tremor and roughness may be similar however roughness occurs at higher modulation frequencies (~20–40 Hz).[29,30]

Contrary to our third hypothesis, no systematic effect of SNR on perceived tremor severity was observed. Although it is possible that SNR does not directly influence the perceived severity of tremor, these results are constrained by the limited number of stimuli and range of SNR evaluated in this experiment. Furthermore, it is possible that SNR covaried with other factors (eg, spectral slope, vowel formants, formant bandwidths), which indirectly affected the perception of tremor. Perceptual relevance of these acoustic features warrants further investigation.

The present study investigated rate ($f_{f0m}$) and extent of $f_0$ modulations ($d_{f0m}$) in patients with vocal tremor. Previous research has shown that changes in $f_0$ modulations can also give rise to modulations in amplitude of a signal.[38] As the frequency of a specific harmonic changes; its distance to the closest formant varies resulting in a change in amplitude. Such interactions were not explicitly evaluated in the current experiment but should be directly explored in another study. As amplitude modulation appears to have the greatest influence on perception of tremor, further investigation of the effects of BOTOX on modulation amplitude need to be explored.

A possible utility of the research described here is to generate a predictive function that gives clinicians and researchers the tools to understand how various treatment approaches might affect perceptual outcomes for individual patients. Successful development of such tools requires a series of perceptual experiments, followed by the development of necessary models or predictive functions. However, the use of a rating scale tasks to gauge changes in perception limits such efforts at this time. This is because rating scale data is biased by several extraneous factors, such as range effects, frequency effects, and bow effects. These factors make the resulting data stimulus-set dependent, and make it difficult to compare findings across experiments, stimuli, or listeners (eg, *see* Shrivastav et al,[37] 2005 and Patel et al,[39] 2010, for more about these issues). To address these issues, perceptual experiments on vocal tremor may also benefit from using different psychophysical tasks, such as the matching task described by Patel et al (2010) for voice quality judgments.[39] Unfortunately, limitations with rating scale data prevent formal modeling of the acoustic-to-perceptual relationships and the nonlinear nature of these findings should be considered as preliminary at this time.

## CONCLUSIONS

The perception of tremor severity for steady-state tremor results from a complex interaction of multiple acoustic cues with $d_{f0m}$

acting as the primary acoustic cue. Other variables have a smaller effect on perceived severity of tremor. This research has implications in assessment and medical or behavioral management of vocal tremor. Regarding assessment, ability to detect voice tremor will depend primarily on $d_{f0m}$ that may affect the degree to which tremor can be perceived in the voice.

## Acknowledgments

We would like to thank David Eddins for designing the experiment in *Sykofizx* software and the two anonymous reviewers for their valuable comments and suggestions.

## REFERENCES

1. Elble RJ. Central mechanisms of tremor. *J Clin Neurophysiol*. 1996;3:133–144.
2. Elble RJ, Koller WC. The diagnosis and pathophysiology of essential tremor. In: *Tremor.* Baltimore, MD: Johns Hopkins University Press; 1990:54–89.
3. Findley LJ, Gretsty MA. Head, facial and voice tremor. In: Jankovic J, Tolosa E, eds. *Advances in Neurology.* New York, NY: Raven Press; 1988:239–253.
4. 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.
5. 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.
6. Dromey C, Warrick P, Irish JC. The influence of pitch and loudness changes on the acoustics of vocal tremor. *J Speech Lang Hear Res*. 2002;45:879–890.
7. Dromey C. Documenting outcomes in vocal tremor. *Voice Voice Disord*. 2004;14:6–8.
8. Shao J, MacCallum JK, Zhang Y, Sprecher A, Jiang JJ. Acoustic analysis of the tremulous voice: assessing the utility of the correlation dimension and perturbation parameters. *J Commun Disord*. 2010;43:35–44.
9. Koller W, Graner D, Mlcoch A. Essential voice tremor: treatment with propranolol. *Neurology.* 1985;35:106–108.
10. Koller WC. Long-acting propranolol in essential tremor. *Neurology.* 1985; 35:108–110.
11. Hartman DE, Vishwanat B. Spastic dysphonia and essential (voice) tremor treated with primidone. *Arch Otolaryngol*. 1984;110:394–397.
12. Muenter MD, Daube JR, Caviness JN, Miller PM. Treatment of essential tremor with methazolamide. *Mayo Clini Proc*. 1991;66:991–997.
13. Busenbark K, Ramig L, Dromey C, Koller WC. Methazolamide for essential voice tremor. *Neurology.* 1996;47:1331–1332.
14. Massey EW, Paulson GW. Essential vocal tremor: clinical characteristics and response to therapy. *South Med J*. 1985;78:316–317.
15. Warrick P, Dromey C, Irish J, Durkin L. The treatment of essential voice tremor with botulinum toxin A: a longitudinal case report. *J Voice*. 2000; 14:410–421.
16. Hertegard S, Granquist S, Lindestad PA. Botulinum toxin injections for essential voice tremor. *Ann Otol Rhinol Laryngol*. 2000;109:204–209.
17. Adler CH, Bansberg SF, Hentz JG, et al. Botulinum toxin type A for treating voice tremor. *Arch Neurol*. 2004;61:1416–1420.
18. Brown JR, Simonson J. Organic voice tremor: a tremor of phonation. *Neurology.* 1963;13:520–523.
19. Hartman DE, Overholt S, Vishwanath B. A case of vocal cord nodules masking essential (voice) tremor. *Arch Otolaryngol*. 1983;108:52–53.
20. Ludlow CL, Bassich CJ, Connor NP, Coultor DC. Phonatory characteristics of vocal fold tremor. *J Phonetics*. 1986;14:509–516.
21. Ramig LA, Shipp T. Comparative measures of vocal tremor and vocal vibrato. *J Voice*. 1987;1:162–167.
22. Winholtz WS, Ramig LO. Vocal tremor analysis with the vocal demodulator. *J Speech Hear Res*. 1992;35:562–573.
23. Horii Y. Acoustic analysis of vocal vibrato: a theoretical interpretation of data. *J Voice*. 1989;3:36–43.
24. Kreiman J, Gabelman B, Gerratt BR. Perception of vocal tremor. *J Speech Lang Hear Res*. 2003;46:203–214.
25. Shrivastav R. The use of an auditory model in predicting perceptual ratings of breathy voice quality. *J Voice*. 2003;17:502–512.
26. Shrivastav R, Sapienza C. Objective measures of breathy voice quality obtained using an auditory model. *J Acoust Soc Am*. 2003;114:2217–2224.
27. Shrivastav R, Sapienza C. Some difference limens for the perception of breathiness. *J Acoust Soc Am*. 2006;120:416–423.
28. Shrivastav R, Eddins DA, Patel S, Cummings S. A rough task: defining a measure for rough voice quality. *J Acoust Soc Am*. 2008;124:2495.
29. Shrivastav R, Eddins DA. Psychometric functions for rough voice quality. *J Acoust Soc Am*. 2010;127:2021.
30. Shrivastav R, Eddins DA. The influence of amplitude modulation frequency on perceived roughness of vowels. *J Acoust Soc Am*. 2011;129:2661.
31. Zwicker E, Fastl H. *Psychoacoustics: Facts and Models*. New York, NY: Springer-Verlag; 1990. 103–132.
32. Kay Elemetrics Corporation. *Disordered Voice Database, Model 4337*. Lincoln Park, NJ: Kay Elemetrics Corporation; 1994.
33. Milenkovic P. *TF32* [Computer software]. Madison, WI: Author; 2001.
34. Kawahara H. Speech representation and transformation using adaptive interpolation of weighted spectrum: vocoder revisited. *Acoust Speech Signal Process*. 1997;2:1303–1306.
35. Baken RJ, Orlikoff RF. *Clinical Measurement of Speech and Voice*. San Diego, CA: Singular Publishing Group; 1993.
36. American National Standards Institute. *Methods for Manual Pure-Tone Threshold Audiometry.* (ANSI S3.21-2004). New York, NY: American National Standards Institute; 2004.
37. Shrivastav R, Sapienza CM, Nandur V. Application of psychometric theory to the measurement of voice quality using rating scales. *J Speech Lang Hear Res*. 2005;48:323–335.
38. Sundberg J. Acoustic and psychoacoustic aspects of vocal vibrato. In: Dejonckere PH, Hirano M, Sundberg J, eds. *Vibrato*. San Diego, CA: Singular; 1995:35–62.
39. Patel S, Shrivastav R, Eddins DA. Perceptual distances of breathy voice quality: a comparison of psychophysical methods. *J Voice*. 2010;24:168–177.