Perception of Vibrato by Professional Voice Teachers: A Pilot Study

[*]Joshua D. Glasner, and [†]John Nix, *Doylestown, Pennsylvania, and †San Antonio, Texas

Summary: Objectives. Voice teachers and clinicians often use vibrato characteristics as auditory-perceptual cues for giving functional instructions. Historical texts also point to the use of vibrato characteristics as a diagnostic tool as far back as the 18th century. This study investigates the relationship between vibrato rate and vibrato extent, and the auditory-perceptual ability of professional voice teachers to assess the vibrato rate of a synthesized singing voice.

Methods. Thirty-five professional voice teachers completed a Visual Sort and Rate task with 20 synthesized singing samples that systematically varied in both vibrato rate (V_{rate}) and vibrato extent (V_{ext}). Two generalized linear mixed effects models and one linear model were designed.

Results. Both V_{rate} and V_{ext} were found to be significantly associated with the perception of V_{rate} as well as with the accuracy of perceptual judgments. Neither listener age, nor amount of teaching experience were significant predictors of perceptual accuracy. Inter-rater agreement for the entire sample set was moderate and intra-rater reliability for samples with identical V_{rate} (but differing V_{ext}) was found to be negligible.

Conclusions. While professional voice teachers have a skilled ability to discern nuanced auditory-perceptual vocal characteristics such as vibrato, in this study, samples with an identical V_{rate}, but for which the V_{ext} differed seemed to demonstrate a high amount of listener rating error and variability. As such, it seems that both V_{rate} and V_{ext} may play a role in the perception of vibrato rate.

Key Words: Singing voice–Vibrato–Auditory-perception–Voice pedagogy–Frequency modulation.

INTRODUCTION

The perception of vibrato characteristics is of particular interest to and holds practical value for voice pedagogues due to its relationship to voice production and performance practice. To date, the literature provides evidence for the physiological, neurological, and acoustical bases of vibrato.[1,2] Vibrato has been reported to depend upon synergies (ie, equilibrium) between the postural, respiratory, phonatory, resonatory, articulatory, and neurological systems.[3–5] It often serves as a means for teachers and clinicians to monitor overall function.[6,7] Furthermore, voice teachers monitor vibrato characteristics to ensure that singers perform within accepted boundaries of specific musical genres, such as earlier periods of Western Classical style (Medieval, Renaissance, Baroque), Contemporary Commercial Music (CCM) and musical theater, and non-Western traditions such as Peking Opera or traditional Hindustani music.[8–11] As such, it is of specific interest, and the objective of this present study, to investigate the perception of vibrato rate by experienced singing voice teachers.

Vibrato

Vibrato in singing is a periodic oscillation of frequency, amplitude, timbre, subglottic pressure, closed quotient, and formant frequencies. Adjustments in the frequency vibrato extent and rate may impact the perception of sung tones. Frequency vibrato extent describes how far above and below the mean frequency that the fundamental frequency (f_o) and all its harmonic components sweep each cycle.[12,13] It is measured as a percentage or in cents, with 6% (100 cents) representing a musical half step.[14] Typical values in Western Classical singing are approximately ± 50 cents, or a quarter tone,[2,15] though the literature is conflicted as to how it reports and measures vibrato extent values.[16] Additionally, recent inquiry suggests that the vibrato extent of modern-day professional Western operatic singers may exceed historically reported norms.[17,18] Frequency vibrato rate refers to the number of frequency oscillations that the fundamental frequency (f_o) and all its harmonic components sweep per second and is measured in Hertz (Hz). Typical values for vibrato rate in Western Classical singing range from 4.5 Hz to 6.5 Hz.[15,19] Reported values for frequency vibrato rate have been shown to have decreased over the past century.[17,18,20]

Perception of Vibrato

Some literature outside of music and singing voice science, such as those related to the perception of vocal tremor, are relevant to this discussion given the possible relationship between vocal vibrato and essential vocal tremor.[1] Kreiman et al[21] reported that differences in tremor rates were most easily perceived when the tremor was sinusoidal and of small amplitude. That is to say, when tremor "extent" was smaller and more regular, participants were able to detect tremor rate differences. The same study also found that differences in tremor rate were easier to perceive when the rate was higher. These findings suggest that differences in the vibrato

Accepted for publication October 26, 2023.
From the *School of Graduate and Professional Studies, Delaware Valley University, 700 E Butler Avenue, Doylestown, Pennsylvania 18901; and the †School of Music, University of Texas at San Antonio, One UTSA Circle, San Antonio, Texas 78249.
Address correspondence and reprint requests to Joshua D. Glasner, School of Graduate and Professional Studies, Delaware Valley University, 700 E Butler Avenue, Doylestown, PA 18901. E-mail: Joshua.glasner@delval.edu
Journal of Voice, Vol xx, No xx, pp. xxx–xxx
0892-1997

https://doi.org/10.1016/j.jvoice.2023.10.032

rate of singing samples might be perceived most accurately in samples with a low vibrato jitter/shimmer (ie, perturbations in cycle-to-cycle vibrato rate and extent, respectively), a smaller frequency extent, and a higher rate. They also indicate a relationship between an individual's auditory-perceptual acuity and a possible psychoacoustic interaction between vibrato rate and vibrato extent—a relationship that has been the subject of recent study related to essential voice tremor severity.[22]

Several studies have examined perceptual aspects of vibrato in vocal and instrumental performance. One such study developed rules for controlling vibrato characteristics in synthesized samples of string instruments. The researcher found vibrato rate to be of greater import to the just-noticeable difference (JND) than vibrato extent.[23] Furthermore, the JND for detecting vibrato rate changes was reported to be approximately 6%.

The literature is robust when discussing the JND of various aspects of musical performance.[24–27] The JND for frequency perception is particularly relevant to the topic of vibrato extent and rate. For pure tones (ie, sinusoidal steady-state signals), the JND of pitch—or the frequency difference limen—varies with f_o, sound intensity level, and duration.[25,28] Studies in speech science point to 500 ms being the durational threshold at which expert listeners may judge a vocal tremor as being moderate to severe.[29] The literature also confirms that vocal tremor is more easily perceived during sustained phonation than in connected speech.[30]

Vibrato characteristics also seem to influence experts' perception of pitch. In a study comparing operatic singers and early music performers, expert listeners were found to assign pitch to a sung sample correctly when the frequency vibrato extent value was smaller.[31] More broadly, the literature suggests that vibrato does seem helpful when ascertaining the pitch of a sung tone, though, of course, the f_o is known to be perceived as the geometric mean of a vibrato note's extreme oscillating frequencies.[32] Furthermore, d'Allesandro and Castellengo[33] demonstrated that the overall f_o oscillation pattern influences pitch perception. Of particular relevance to the present study is their finding that the "separation/fusion of pitch perception" is related to and influenced by both vibrato extent and vibrato rate.

In addition to improving the perceptual accuracy of intonation differences, vibrato has been shown to influence listeners' ability to discern the number of voices in an audio sample.[34,35] Erickson et al analyzed the perception of untrained listeners using synthesized stimuli to examine the role of vibrato—among other factors—in (a) the perceived difficulty of discerning the number of voices performing and (b) the ability of listeners to perceive differences in intonation. They found that vibrato and formant patterns in the stimuli interacted with the listeners' ability to hear two simultaneous voices.[35] Furthermore, their work agrees with Daffern et al[31] and supports the concept that listeners perceive a sample without vibrato as more pitch-different than samples with vibrato. That is to say, listeners determine pitch more accurately when the vibrato extent is smaller.

Listeners also appear to be sensitive to differences in cyclic variations in loudness—termed amplitude vibrato (or loudness vibrato). While frequency vibrato is both extensively studied and reported to be influential to listeners,[36] there is some literature that suggests that cyclic spectral variations play a fundamental role in the perception of vibrato.[37] As harmonics of the fundamental frequency sweep through vocal tract resonances, amplitude and timbral variations occur, sometimes in phase with; sometimes 180 degrees out of phase with; and sometimes at twice the rate of the frequency vibrato—depending upon the location of resonances relative to the harmonics.[36,38] However, this sensitivity "has its limits," as more recent research indicates that listeners cannot detect fluctuations of loudness or spectral centroid location below approximately 15%.[39] The same authors also reported that "proportional fluctuations in loudness are perceived as being about 1.3 times deeper [ie, larger in extent] than proportional fluctuations in spectral centroid;[39]" suggesting that listeners are more sensitive to loudness variations than to those of timbre.

In addition to discussions of the perceptual thresholds and sensitivity of listeners to differences in vibrato characteristics, it is also of value to study the preference of professional musicians. Trained singers who have primarily choral experience have been found to prefer a narrow vibrato extent.[40] Furthermore, trained voice professionals have been found to judge a tone to be without vibrato when the peak-to-peak extent falls below 27 cents.[41] The same study also noted that the measured vibrato rate significantly affects where this extent threshold occurs. This observation may be explained by literature that suggests that the perception of vibrato extent may be tied to the rate of frequency changes (RFC) (measured in cents per second).[42] RFC varies with a change in frequency vibrato rate or extent. In other words, the ability of listeners to identify or perceive differences in vibrato characteristics may be a function of proportional changes between vibrato rate and vibrato extent, and those perceptual distinctions may be temporally variant.

Of particular interest to the authors of this study is the variable use of expert and amateur listeners in studies related to vibrato and timbre perception. Contrary to the results reported by Wooding and Nix,[41] Reddy and Subramanian[43] found that there were significant differences between the experienced and inexperienced groups of judges. More specifically, singer (ie, experienced) judges perceived vibrato characteristics with more detail than the inexperienced judges. Yet, other studies suggest that measurements of vibrato extent and rate are only moderately correlated with the judgments of expert listeners.[44] It seems that the use of expert or non-expert listeners should be determined by the goals of individual studies. As such, those studies which aim to inform pedagogical or clinical practice—such as the study described in this paper—should consider either training listeners as a component of the research method or studying the evaluations of expert listeners.

Lastly, while it might be reasonable to argue against the ecological validity of using synthesized voice samples to study the auditory-perceptual abilities of skilled listeners—professional voice teachers in this case—the practice of using synthesized audio samples to study auditory-perceptual characteristics of voicing has historical precedence in the literature.[22,29,33,34,40–42] Furthermore, studies suggest that the "context in which a quality judgment is made…" is primarily responsible for listener agreement.[45] In the context of the current study, it was assumed that expert listeners would be likely to classify samples at a "…specific, subordinate level of abstraction…"[46] and that they may be unduly influenced by stimuli if there were additional timbral or phonatory cues. This inquiry specifically chose to prioritize synthesized samples for which phonatory and timbral factors were systematically controlled. As such, this study aimed to use synthesized audio samples to carefully control for psychoacoustic factors that may influence an individual's auditory-perceptual judgment. The use of synthesized samples will be outlined further in Methods.

In summary, past literature suggests an intriguing connection between essential vocal tremor and vocal vibrato. Historical voice pedagogy texts also point to the importance of vocal vibrato as a diagnostic tool for vocal function and as an indicator of performance practice. Studies in both speech-language pathology and voice pedagogy impress the importance of auditory-perceptual training in clinical and performance practice—a "gold standard" for clinicians and singing teachers. As such, it is the objective of this study to understand the ability of experienced voice teachers to discern characteristics related to vocal vibrato. In so doing, it may be possible to develop a framework which systematically and efficiently trains the auditory-perceptual skills of future clinicians and voice teachers.

Hypotheses

The purpose of this study was to determine the accuracy and reliability of voice professionals in making judgments about the frequency vibrato rate of synthesized singing voice samples. Additionally, this paper investigated the relationship between demographic factors, such as age and years of teaching experience, and rating accuracy. We hypothesized (1) that vibrato rate and vibrato extent both would be significant predictors of the perception of vibrato rate, (2) that both vibrato rate and vibrato extent would be significant predictors of perceptual accuracy, and (3) that demographic factors would not be significant predictors of perceptual accuracy.

METHODS

Permission for the study protocol was obtained from the University of Texas at San Antonio IRB (UTSA IRB FY20-21-247). Participants were recruited via messages on the ASHA Special Interest Group 3 list-serv, on social media forums for singing teachers, and by direct email. Direct recruitment emails were sent to 162 singing teachers, with specific attention toward recruiting a gender and ethnically diverse participant population. The qualifications of the study included the following self-reported criteria:

- Singing teachers residing in the U.S.
- Minimum age of 22 years.
- Minimum education of a completed undergraduate degree in music.
- Intact hearing (self-reported, with corrective hearing aids if necessary).
- Access to a computer and headphones for playing audio files.

Participants

Sixty-seven teachers responded to recruitment efforts, and 35 (52%) completed all aspects of the perceptual protocol. Demographic information about the participants can be found in Tables 1 and 2. Note that of the 35 participants who completed the perceptual protocol, only 30 participants also fully completed the questionnaire.

Equipment

Twenty synthesized singing samples were created using *Madde* (Tolvan Data, Stockholm, Sweden) on the pitch C5 (~523 Hz). The vibrato extent and rate measurements for each sample are found in Table 3 and were chosen so as to approximate historically reported norms as well as values outside of that range. The vowel used for all samples was [a] with the following supraglottal vocal tract resonance frequencies: F_1 = 816 Hz, F_2 = 1346 Hz, F_3 = 3031 Hz, F_4 = 3984 Hz. The duration of each sample was 5 seconds, which was chosen to allow listeners enough time to notice frequency-related changes given that duration is a key factor in the JND of frequency. Of course, given that participants were permitted unlimited opportunities to listen to stimuli, the duration of each sample was of less import than controlling acoustic parameters. As such, the default *Madde* parameters were maintained for all synthesized samples (source spectrum tilt = −6 dB/octave; flutter = 1%; Q1-Q6 = 10).

Each sample was normalized to a relative sound level of 68 dB in *Praat*.[48] A gradual intensity modulation was applied to the first and last 0.5 second of each 5-second sample in *Audacity* (v.2.4.2.0) (The Audacity Team) to ensure that participant responses were not influenced by the onset or the offset (ie, attack and decay) of the synthesized samples. Each file was then converted to *mp3* to limit the file size for participants. Note that the sound level of the playback was not controlled outside of ensuring that the original audio files were normalized to 68 dB.

Unrelated to the file conversion, but of import for future protocols is that the authors observed following data collection that the audio samples were bandlimited at 3.5 kHz. Upon further investigation, it seems that this spectral limitation is inherent to the *Madde* application.

Table 1.
Demographic Information (Categorical) for All Participants Who Completed the Questionnaire (n = 30)

Characteristic	Frequency (Relative Frequency)
Gender identity	
Female	20 (66.7%)
Male	9 (30.0%)
Non-binary	1 (3.3%)
Voice type	
Soprano	16 (53.3%)
Mezzo soprano	4 (13.3%)
Tenor	4 (13.3%)
Baritone	4 (13.3%)
Bass	1 (3.3%)
Countertenor	1 (3.3%)
Teaching environment	
Full-time academic and private	14 (46.7%)
Part-time academic and private	7 (23.3%)
Academic only	5 (16.7%)
Private only	4 (13.3%)
Primary style of singing (perform)	
Western classical	23 (76.7%)
Musical theater	3 (10.0%)
Jazz	1 (3.0%)
Opera	1 (3.0%)
Other	2 (5.9%)
Folk, Gospel, Pop, Rock, R&B	0 (0.0%)
Primary style of singing (teach)	
Western classical	19 (63.3%)
Musical theater	6 (20.0%)
Pop	3 (10.0%)
Opera	1 (3.0%)
R&B	1 (3.0%)
Listen to vibrato as an indicator of vocal function	
Always	20 (66.7%)
Sometimes	8 (26.7%)
Seldom	2 (5.9%)
Never	0 (0.0%)

Table 2.
Demographic Information (Numerical) for All Participants Who Completed the Questionnaire (n = 30)

	Mean	Min	Max	SD
Age	46.4	23.0	72.0	13.6
Total teaching experience	21.0	2.0	45.0	12.5
Full-time teaching experience	14.6	0.0	42.0	11.2

Table 3.
List of the Sample Codes in Addition to the True Rank (Based on Vibrato Rate), Vibrato Rate, and Vibrato Extent for Each Token

Sample Code	True Rank	V_{rate} (Hz)	V_{ext} (cents)
T	1	4.8	100
L	1	4.8	100
C	3	5	100
G	4	5.2	100
W	4	5.2	100
X	6	5.4	100
H	7	5.5	100
I	7	5.5	100
Q	7	5.5	100
DD	7	5.5	100
N	7	5.5	60
J	7	5.5	80
A	7	5.5	200
D	7	5.5	180
B	7	5.5	160
E	7	5.5	120
P	17	5.8	100
AA	18	6	100
R	18	6	100
BB	20	6.2	100

Protocol

Each participant was emailed a link to a demographic questionnaire as well as one of two PowerPoint slides (v.16.64) (Microsoft Corporation, Redmond, WA). The two versions of the slide included 20 randomly sorted voice samples embedded within the slide. The protocol used for this study followed a variation of the Visual Sort and Rate (VSR) method suggested by Granqvist.[47] Participants opened the PowerPoint slide, clicked on the icon for each sample, listened to it through headphones as many times as necessary, and placed the icon along an x-axis (24.66 cm) based on the rate (ie, speed) of the vibrato. Figure 1 shows a blank example of the slide. Participants were naïve to the fact that some stimuli were identical in rate (see Figure 1 for the participant instructions). As participants listened to more samples, they rearranged the icons until they were satisfied with their sorting and rating. The completed slide was then saved and emailed back to the investigators for

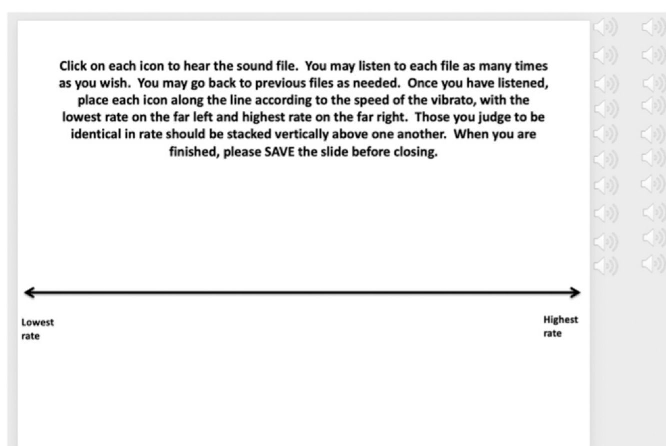

FIGURE 1. Image of the blank Powerpoint slide. Each participant was given to use, to sort, and rate the synthesized singing samples.

analysis. Note that there was no prescribed time limit to complete the VSR task following Granqvist.[47]

Analysis

Analog Data

While Granqvist[47] suggested the use of a free application, *Visor* (Tolvan Data, Stockholm, Sweden), for VSR tasks, the present study prioritized years of teaching experience rather than technological aptitude or facility. As such, Adobe Acrobat Pro DC (v.2022.001.20112) (Adobe, Inc., San Jose, CA) was used to analyze the analog data in each PowerPoint slide.

Each slide was first converted to PDF format. Distance measurements were collected using the "Measuring Tool" plugin feature, which was standardized to the document size (27.94 cm). To ensure consistency, the terminal right end of the measurement tool was placed at the beginning of each icon.

Statistical Analysis

Statistical analysis was performed using R 4.1.2 (R Core Team, Vienna, Austria) and RStudio (v.1.5.46) (RStudio, Boston, MA). The lme4 package was used for all modeling procedures.[49] The irr package was used to estimate agreement between participants and reliability within participants for samples with an identical vibrato rate value.[50] The confidence level for statistical analysis was set at 95% ($\alpha = 0.05$).

This study aimed to understand the predictive relationship between vibrato rate and listener perception. As such, two linear mixed effects models were designed. The first model included the distance on the x-axis (d) as the dependent variable, participant as a random effect, and vibrato rate (V_{rate}) and vibrato extent (V_{ext}) as fixed effects. The second model used the absolute difference between the true rank and the rank determined by each participant for a sample ($|\Delta_{rank}|$) as the dependent variable, and, otherwise, was identical to the first model. The residuals for both models were found to violate the assumption of normality using the Shapiro-Wilk test. As such, both models were redesigned as generalized linear mixed effects models (GLMM) fit by maximum likelihood (Laplace approximation).

For the first model, d, V_{rate}, and V_{ext} were scaled and centered. They were subsequently transformed by adding a constant to the scaled variables to allow for GLMM analysis using a Gamma family distribution (ie, no non-positive values). The second model followed the same process for the variables $|\Delta_{rank}|$, V_{rate}, and V_{ext}.

A final linear model was designed with the sum of the absolute rank differences for each participant ($\Sigma|\Delta_{rank}|$) as a dependent variable and with the following independent variables: age, teaching experience, and full-time teaching experience. No random effects were considered as each participant only had one value for $\Sigma|\Delta_{rank}|$.

Two intraclass correlation coefficient models (ICC) were calculated to estimate listener agreement and reliability. The estimate of inter-rater agreement was calculated based on a single-rating, absolute agreement, two-way random effects model. The estimate of intra-rater reliability was calculated based on a single-rating, absolute-agreement, two-way mixed effects model.

RESULTS

The present study follows in part the VSR method as suggested by Granqvist[47] for assessing perceptual acuity of voice samples in recordings. Listeners were professional voice teachers, for whom descriptive statistics can be found in Tables 1 and 2. The participants were given one of two prepared slides and were asked to sort and rate audio samples of a synthesized singing voice along an x-axis from lowest vibrato rate to highest vibrato rate. Submissions were analyzed according to the distance of each sample from the origin on the x-axis (0.00–24.66 cm) and according to their difference in rank compared to the true rank of the audio samples (Table 2). The vibrato rate and vibrato extent of each synthesized sample were known to the authors, but were not revealed to the participants at any point.

Perception of Vibrato Rate

For the purpose of this study, the placement of an audio sample icon along the x-axis and, thus, the distance measurement for each sample is considered equivalent to a listener's perception of vibrato rate. A GLMM fit by maximum likelihood (Laplace Approximation) was designed with d as the dependent variable, participant as a random effect, and V_{rate} and V_{ext} as fixed effects. It was controlled with a BOBYQA (Bound Optimization BY Quadratic Approximation) optimizer.[51] All variables except for the random effect were scaled, centered, and transformed with a constant. The model converged and showed significant negative associations of V_{rate} and V_{ext} with d (Table 4).

Accuracy

Rank accuracy was calculated by taking the absolute difference between the participant's ranking of an audio

Table 4.
Summary of the Fixed Effects for the GLMM for the Perception of Vibrato Rate

Fixed Effects:	Estimate	SE	*t*-value	Pr (> |z|)	
V_{rate}	−0.083	0.003	−29.203	0.000	*
V_{ext}	−0.011	0.003	−3.561	0.000	*

The perception of vibrato rate was measured as the distance value along the x-axis (d) for each audio sample within each listener's completed slide.
Abbreviation: SE, standard error.
* Indicates values that were found to be significant at the 95% confidence level.

Table 5.
Summary of the Fixed Effects for the GLMM for the Rank Accuracy

Fixed Effects	Estimate	SE	*t*-value	Pr (>|z|)	
V_{rate}	−0.009	0.004	−2.456	0.014	*
V_{ext}	−0.014	0.003	−4.205	0.000	*

The rank accuracy was measured as the absolute difference between the true rank and the listener's rank for each audio sample (|Δrank|).
Abbreviation: SE, standard error.
* Indicates values that were found to be significant at the 95% confidence level.

sample and the true rank. A GLMM fit by maximum likelihood (Laplace Approximation) was designed with $|\Delta_{rank}|$ as the dependent variable, participant as a random effect, and V_{rate} and V_{ext} as fixed effects. It was controlled with a BOBYQA optimizer. V_{rate} and V_{ext} were scaled, centered, and transformed with a constant. The model converged and showed significant negative associations of V_{rate} and V_{ext} with $|\Delta_{rank}|$ (Table 5).

Demographics and Total Error

A linear model was designed to study the predictive ability of age and teaching experience on total ranking error of professional listeners ($\sum|\Delta_{rank}|$). Note that this model only included participants who completed both the perceptual task and the questionnaire (n = 30). The model was not found to explain a significant amount of variance ($R^2_{adj} = -0.047$, $F_{3,26} = 0.5651$, $P = 0.643$). The coefficients for the independent variables (age, teaching experience, and full-time teaching experience) as well as their respective *P*-values can be found in Table 6. No demographic variable was found to be a significant predictor of a listener's total ranking error.

Listener Agreement

The ICC (2,1) was calculated to test for two-way agreement between participants for *d* (perception of vibrato rate). Two participants each did not provide a rating for one sample. Those samples (DD and X) were automatically omitted from the dataset for the calculation of ICC in a listwise way. Agreement between listeners for 18 samples was estimated to be 60.5% (ICC = 0.605, $F_{(17,445)}$ = 64.2, $P < 0.000$). The confidence interval for ICC population values was found to be 0.456–0.778.

Table 6.
Summary of the Linear Model for Demographic Information as a Predictor of Total Ranking Error ($\sum|\Delta_{rank}|$)

Model Coefficients:	β Estimate	SE	*t*-value	Pr (>|*t*|)
Age	0.590	0.815	0.724	0.476
Teaching experience	−0.295	1.095	−0.270	0.789
Full-Time (FT) teaching experience	0.064	0.841	0.076	0.940

The total ranking error rank was measured as the sum of the absolute differences between the true rank and the listener's rank for each audio sample. No independent variable was found to be a significant predictor of total ranking error.
Abbreviations: FT, full-time; SE, standard error.

Listener Reliability

Intra-rater reliability was estimated for listeners' rating of samples for which the vibrato rate was identical (5.5 Hz). This portion of the dataset included 10 samples. One was incomplete and was automatically omitted in a listwise way. Reliability within listeners for nine samples was estimated to be 0% (ICC = 0.000, $F_{(8272)}$ = 1.00, $P = 0.440$). The confidence interval for intra-rater reliability was found to be −0.014 to 0.053.

DISCUSSION

This paper aims to investigate voice professionals' ability to accurately perceive vibrato rate. To do so, it presented 35 professional voice teachers with 20 synthesized singing voice audio samples for which vibrato parameters—vibrato rate and vibrato extent—varied systematically. While having singing teachers perceptually evaluate synthetic samples of modulated phonation is not entirely reflective of their professional activities (ie, training human artists), it is one of the most controlled means for perceptual investigations available to present-day singing voice researchers.

The results of this inquiry suggest the following: (1) that vibrato rate and vibrato extent both may be factors in the perception of vibrato rate by professional voice teachers, (2) that alterations of both vibrato rate and vibrato extent may result in difficulty correctly ordering audio samples based upon vibrato rate (ie, accurately perceiving vibrato rate differences), and (3) that demographic factors such as age and teaching experience do not seem to significantly predict the perceptual acuity of professional voice teachers in regard to assessing vibrato rate. Agreement between listeners was estimated to be approximately 60.5% and intra-rater reliability for samples with identical vibrato rate—but differing vibrato extent—was negligible. These findings further emphasize the complex perceptual nature of vocal vibrato. Furthermore, they indicate that, while professional voice teachers do seem to have a developed ability to discern the vibrato rate of audio samples, vibrato extent may influence even a professional's ability to distinguish differences in vibrato rate.

Perception of Vibrato

Results related to the VSR scale (Figure 2) suggest that both vibrato rate and vibrato extent influence professional voice teachers' ability to perceive vibrato rate in synthesized singing samples. More specifically, as vibrato extent became larger, or wider, participants tended to rate audio samples as having a slower vibrato rate. This finding seems

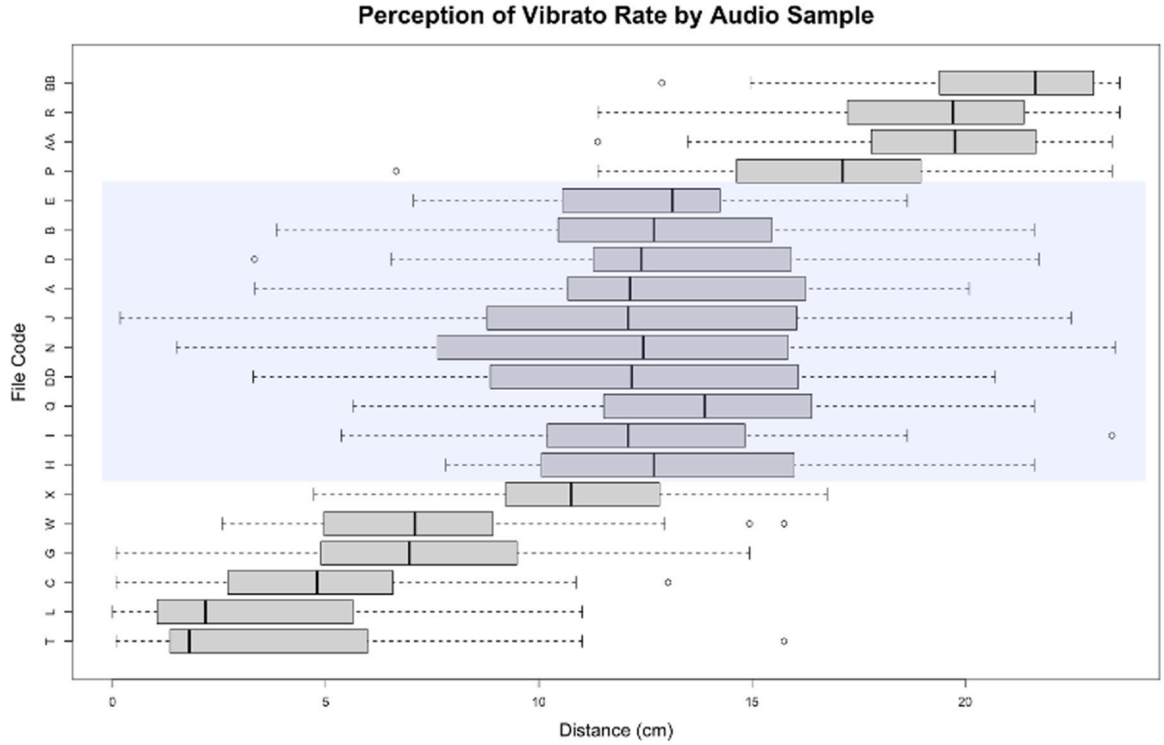

FIGURE 2. A boxplot of each audio sample and the range of distance values assigned by listeners. The blue box highlights audio samples for which the vibrato rate was 5.5 Hz. Note the variability in the highlighted samples compared to those for which the vibrato extent was constant (100 cents).

to support the findings of Vatti et al,[42] who suggested that a relationship exists between a listener's judgment of vibrato rate differences and the RFC in a singing sample—a metric that represents both the vibrato rate and the vibrato extent of a sample.

Curiously, there was a negative association between vibrato rate and the distance along the x-axis. This abnormality might be a function of the number of samples with an identical vibrato rate (5.5 Hz) (Table 2). The selection of synthesized samples for perceptual studies of the singing voice will be discussed in further detail in Future Research. Nonetheless, based on the results of this study, it seems prudent for clinicians to consider that their ability to discern the vibrato rate of a singer might be related not just to the singer's demonstrated vibrato rate, but also to the singer's simultaneously occurring vibrato extent.

Accuracy

Both vibrato rate and vibrato extent seem to affect the ability of professional voice teachers to accurately sort and rate synthesized singing samples based on vibrato rate. The results from this study suggest that listeners were less accurate when the vibrato rate was constant and when the vibrato extent was independently altered. Based upon the models presented above, it might be tempting, though errant, to suggest that a smaller vibrato extent leads to a generalized and decreased ability to ascertain the vibrato rate of a singer. Instead, the direction of the association between vibrato extent and perceptual accuracy can be thought of as an artifact of the present research method, though the presence of a significant association between vibrato extent and the perception of vibrato rate warrant further investigation.

Consider Figure 3 which uses RFC to visualize the combined relationship of vibrato extent and vibrato rate with the mean ranking error across participants. It also contrasts synthesized samples for which the vibrato extent varied with those for which vibrato rate varied. From this visualization, it seems that ranking error was higher on average for audio samples with a vibrato rate of 5.5 Hz (50% of the samples). In other words, listeners seemed to have difficulty accurately sorting samples based upon vibrato rate when the vibrato rate was identical, but the vibrato extent was variable. As such, the results of this study seem to indicate that vibrato rate is not the sole influence on the accurate perception of vibrato rate.

Demographics and Total Error

Anecdotally, the fields of voice performance and voice pedagogy emphasize the relationship between years of experience and auditory-perceptual acuity. Indeed, in all forms of Evidence-Based Practice, including Evidence-Based Voice Pedagogy, clinical or voice teacher expertise is held as an equal part of a triumvirate of values that informs best practice.[52,53] While experience is an important component of voice training, the results of this study suggest

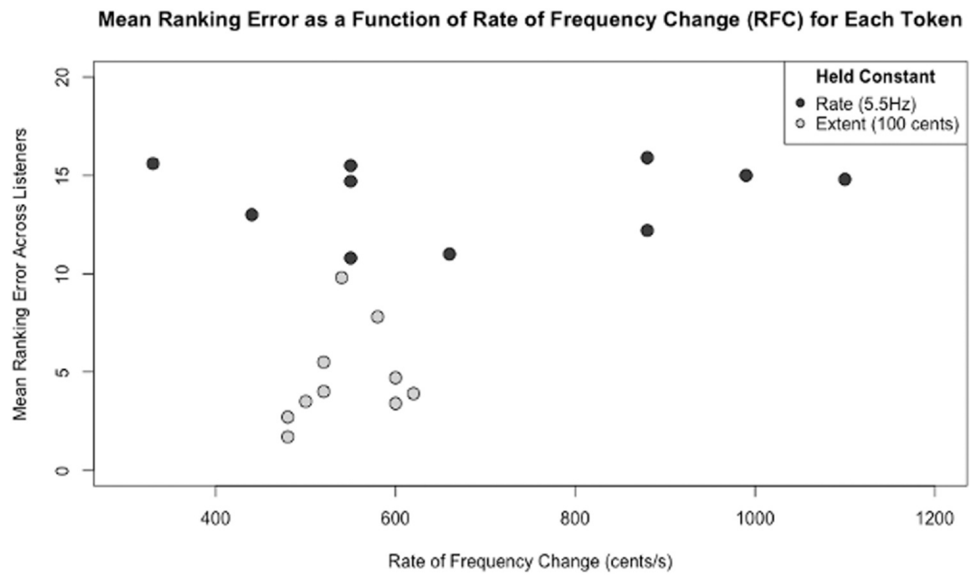

FIGURE 3. Scatterplot of the relationship between mean ranking error across all listeners for each token and the rate of frequency change (RFC) for each token. Gray points represent tokens for which the vibrato extent was 100 cents, but the vibrato rate was variable. Black points represent tokens for which the vibrato rate was 5.5 Hz, but the vibrato extent was variable.

that experience and age alone may not be significant predictors of a listener's ability to accurately determine vibrato rate in singing voice samples. This finding may be a result of having a participant pool with 21.6 years of teaching experience on average, or it may be influenced by a participant pool with a mean age of 46.8 years. It is possible that future studies may determine a threshold of teaching experience that results in high auditory-perceptual acuity; however, it is likely that specific and directed auditory-perceptual training may prove to be more efficacious in helping voice professionals to develop auditory-perceptual judgment of vibrato characteristics.

Listener Agreement

While age and teaching experience did not significantly predict participants' total error in the VSR task, estimated agreement between listeners was found to be poor-to-good based on the estimated confidence interval for the ICC (0.456–0.778).[54] This study also found that intra-rater reliability for the subset of samples with a vibrato rate of 5.5 Hz was negligible. In combination, these two findings suggest that, while professional listeners may have moderate agreement (or poor-to-good) in terms of their perception of vibrato rate, samples that differ in vibrato extent, but not vibrato rate may tend to result in extreme perceptual variability. This finding is in agreement with Wooding and Nix who reported that a narrow vibrato extent can confound listeners' judgment of vibrato and non-vibrato tones.[41]

Limitations

As has been acknowledged throughout this paper, a primary limitation of this study is that the vibrato rate values of synthesized samples were not monotonically increasing. This oversight necessitated a more complicated, though appropriate, statistical analysis and led to less generalizable results. Additionally, though the statistical analysis used in this study was more complex than that planned for in the original research design, many participants attempted to "stack" stimuli icons when they believed the vibrato rate to be identical (Figure 4). This observation suggests that participants overarchingly understood the directions.

Many participants reported the VSR task to be rather onerous and, in some cases, to require over 1 hour to complete. Obviously, a perceptual task that requires 1 hour or more to complete can result in listener fatigue or even participant attrition[47]; however, it is unclear how future studies might both target a participant pool of professional voice teachers without allowing for geographic diversity. As such, controlling for time while allowing for the full benefit of the VSR method—which includes improved listener reliability—might be challenging.

This study prioritized listener experience and geographic diversity so as to analyze the auditory-perceptual abilities of professional voice teachers; however, one difficulty that arose from this choice was the authors' inability to control for playback sound levels. As such, a limitation of this paper is that individual computers may introduce undesirable effects related to signal processing. Additionally, while all participants were asked to document their use of headphones, those technologies are known to introduce varying transfer functions, which may result in unknown filtering effects or distortions.

One additional concern is that the application *Madde* seems to bandlimit synthesized signals at 3.5 kHz. Lester-Smith and

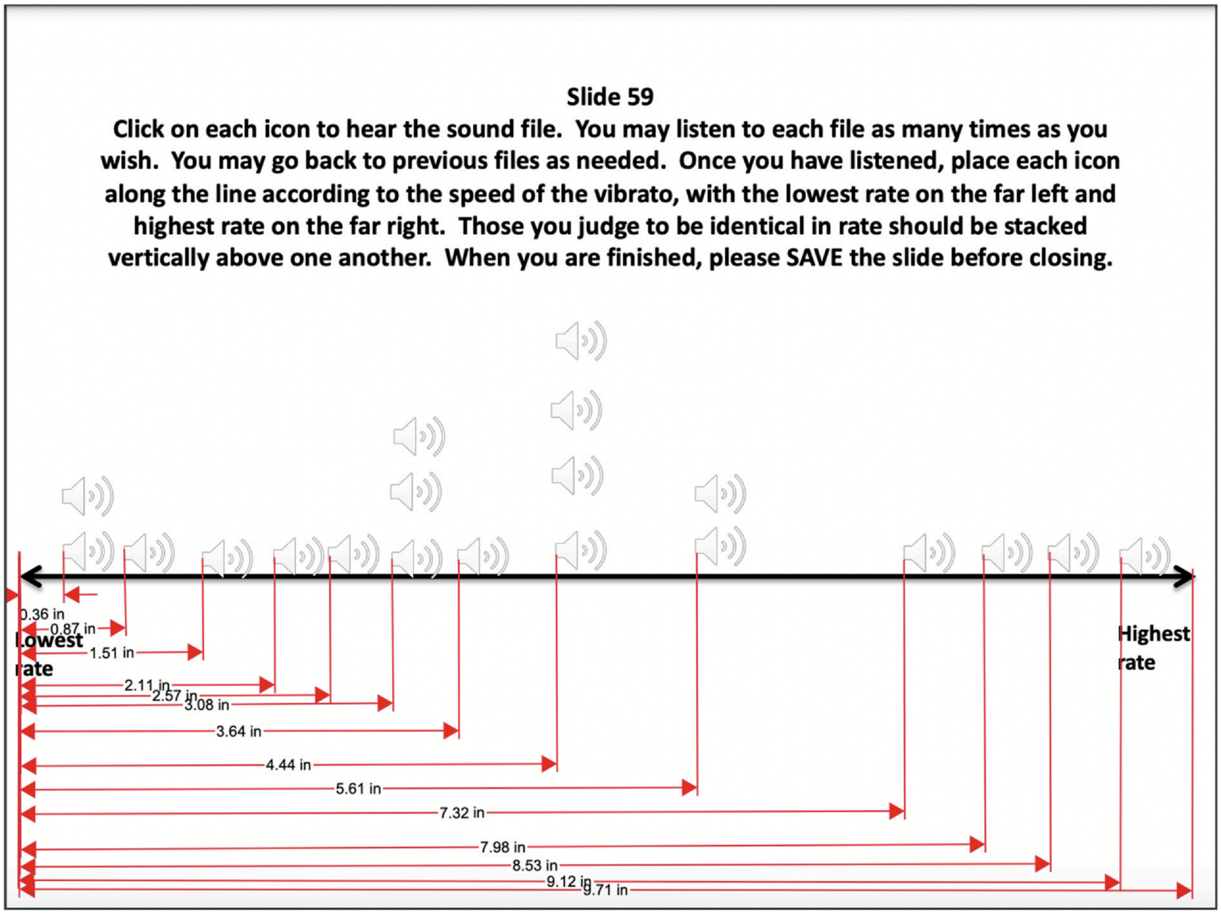

FIGURE 4. Image of a completed Powerpoint slide by a participant. This figure demonstrates the "stacking" mechanism with which many participants completed the VSR task as well as the measuring plug-in used to calculate distance.

Story suggest that spectral energy above 5 kHz may contribute to the perception of frequency modulations in the human voice.[7] Furthermore, Monson et al point to spectral energy above 5 kHz and up to 20 kHz as potentially playing a significant role in speech and voice perception.[55] It is also possible that normalizing the SPL of stimuli may have made it more difficult for listeners to discriminate between tones given that SPL may vary when vibrato extent causes resonance-harmonic interactions. Likewise, the use of .mp3 conversion may have introduced unwanted acoustic distortions. As such, it is possible that there was a trade-off, of sorts, in the use of easily manipulatable synthesized singing samples and easily transferable files to study the perception of voice professionals.

Just as the bandlimit imposed by *Madde* may have adversely influenced the participants in this pilot study, the use of self-reported "intact hearing" as an inclusion factor may be a limitation of this study. Should participants have had limited hearing capabilities, it is reasonable to expect that their performance for this psychoacoustic protocol may have been negatively impacted. The authors trust the participants of this study. Nonetheless, the use of self-reported "intact hearing" is a limitation of this study and should be controlled in the future if possible.

Future Research

Future studies using the VSR method for perceptual studies related to the singing voice should be certain to randomize the parameters for synthesized samples. This type of planning can be carried out easily with functions in R, Matlab, or even a simple online randomizer depending on one's facility with each tool. Randomized parameters that are the subject of inquiry should also be monotonically increasing so as to facilitate simpler statistical procedures. A primary suggestion from this pilot study relates to method design. For example, this study could have implemented a far simpler statistical analysis method if it had used an equally spaced two-dimensional grid with five possibilities of vibrato rate and five possibilities of vibrato extent (25 samples), as well as five redundant samples to test for intra-rater reliability. Furthermore, it might also be prudent in the future to consider using an application such as *VISOR* to reduce the possibility of manual error while measuring analog data. It would be simple to create a short tutorial video to train even geographically diverse participants or participants who are at risk of being less technologically literate.

In regard to suggestions that are not related to research design, this pilot study posits, in agreement with past literature,[41,42] that there is a complex relationship between the perception of vibrato rate and vibrato extent. Future studies should consider investigating if RFC can adequately capture the relationship between the combination of these two parameters and the perception of vibrato. It would also be relevant to study the difference limen for vibrato rate and vibrato extent, again, with the help of professional voice teachers and voice specialist clinicians.

There is still a pedagogical and scientific need to understand the functional causes of vibrato rate and extent more fully.[2,18,56] While it is helpful to study the perception of professional voice teachers and clinicians, it is necessary to explore how they might facilitate functional change in their clients and to use scientific inquiry to move beyond folk pedagogy.

Finally, given that both vibrato rate and vibrato extent are factors in the accurate perception of vibrato rate and, as such, of the percept of vibrato as a whole, future studies should consider how to train the auditory-perceptual accuracy of future voice teachers and clinicians. Past literature—both scientific and otherwise—has pointed to the benefit of using vibrato as a diagnostic tool when training singers. As such, it may be advisable to train future teachers to distinguish independently between vibrato rate and vibrato extent[57] or, similarly, to aurally identify patterns that are appropriate based upon performance practice or functional training.

CONCLUSION

The results of this study suggest that there is a complex relationship between vibrato rate and vibrato extent in the perception of vibrato. Specifically, they show a significant association between both vibrato extent and vibrato rate with the perception of vibrato rate. Furthermore, while professional voice teachers had moderate agreement in their sorting and rating of vibrato rate, their perceptual accuracy diminished when forced to perform the VSR task for samples with the same vibrato rate, but with different vibrato extent measurements. These results suggest that voice professionals who have experience with nuanced auditory-perceptual judgments may characterize a singer's vibrato characteristics based on multiple factors—a concept that has precedence in the literature.

In conclusion, this study found (1) that both vibrato rate and vibrato extent may be significantly associated with professional voice teachers' ability to perceive vibrato rate, (2) that both vibrato rate and vibrato extent may be significantly associated with perceptual accuracy, and (3) that listener age and teaching experience may not be significant predictors of auditory-perceptual accuracy. The authors underscore the quasi-hypothesis-generating nature of this pilot study and, as such, suggest that these results be interpreted cautiously. A primary suggestion resulting from this paper is that it may be pedagogically relevant to train future voice teachers to accurately distinguish between changes in vibrato rate and vibrato extent.

Declaration of Competing Interest

None.

Acknowledgments

The authors wish to thank P.B. and R.S. for their helpful discussions about this paper and for their encouragement.

References

1. Ramig LA, Shipp T. Comparative measures of vocal tremor and vocal vibrato. *J Voice.* 1987;1:162–167. https://doi.org/10.1016/SO892-1997(87)80040-1.
2. Titze I, Story B, Smith M, et al. A reflex resonance model of vocal vibrato. *J Acoust Soc Am.* 2002;111:2272–2282. https://doi.org/10.1121/1.143495.
3. Doscher B. *The Functional Unity of the Singing Voice*. Lanham, MD: Scarecrow Press 1994.
4. Patinka P, Nix J. Changes in vibrato rate, vibrato extent, and vibrato jitter in soprano voices in response to changes in mouth opening: a pilot study. *J Sing.* 2020;77:213–218.
5. Lester-Smith R, Derrick E, Larson C. Characterization of source-filter interactions in vocal vibrato using a neck-surface vibration sensor: a pilot study. J Voice. In press. https://doi.org/10.1016/j.voice.2021.08.004.
6. Nix J. Shaken, not stirred: practical ideas for addressing vibrato and non-vibrato singing in the studio and choral rehearsal. *J Sing.* 2014;70:411–418.
7. Lester-Smith R, Story B. The effects of physiological adjustments on the perceptual and acoustical characteristics of vibrato as a model of vocal tremor. *J Acoust Soc Am.* 2016;140:3827–3833. https://doi.org/10.1121/1.4967454.
8. Strempel E. The shifting aesthetics of vibrato. *J Sing.* 2006;62:405–411.
9. Becker AS, Watson PJ. The use of vibrato in belt and legit styles of singing in professional female musical theater performers. *J Voice.* 2022;22:1–7. https://doi.org/10.1016/j.jvoice.2022.07.018.
10. Han Q, Zhang R. Acoustic analysis of singing vibrato in traditional Peking opera. *J Voice.* 2017;31:P511.E1–511.E9. https://doi.org/10.1016/j.jvoice.2016.11.016.
11. Radhakrishnan N, Scherer RC, Bandyopadhyay S. Laryngeal dynamics of pedagogical Taan gestures in Indian classical singing. *J Voice.* 2011;25:e139–e147. https://doi.org/10.1016/j.jvoice.2010.03.003.
12. Schutte HK, Miller DG, Svec J. Measurement of formant frequencies and bandwidths in singing. *J Voice.* 1995;9:290–296.
13. Hirano M, Hibi S, Hagino S. Physiological aspects of vibrato. In: Dejonkere PH, Hirano M, Sundberg J, eds. *Vibrato.* Norwich, England: Singular Publishing Group 1995:9–33.
14. Rossing TD. *The Science of Sound*. 2nd ed. Boston, MA: Addison-Wesley 1990.
15. Sundberg J. Acoustic and psychoacoustic aspects of vocal vibrato. *STL-QPSR.* 1994;35:45–68.
16. Nestorova T, Glasner JD. What about extent: examining current vibrato extent metrics. *PAVA InFormant.* 2021;3:3–11.
17. Ferrante I. Vibrato rate and extent in soprano voice: a survey on one century of singing. *J Acoust Soc Am.* 2011;130(3):1683–1688. https://doi.org/10.1121/1.3621017.
18. Glasner JD, Johnson AM. Effects of historical recording technology on vibrato in modern-day opera singers. *J Voice.* 2022;36:464–478. https://doi.org/10.1016/j.jvoice.2020.07.022.

19. Titze I. *Principles of Voice Production*. Iowa City, IA: National Center for Voice and Speech 2000.
20. Crutchfield W. Vocal performance in the nineteenth century: an overview. In: Lawson C, Stowell R, eds. *The Cambridge History of Musical Performance*. Cambridge University Press, 2012:611–642.
21. Kreiman J, Gabelman B, Gerratt BR. Perception of vocal tremor. *J Speech Lang Hear Res*. 2003;46:203–214. https://doi.org/10.1044/1092-4388(2003/106).
22. Maryn Y, Leblans M, Zarowski A, et al. Objective acoustic quantification of perceived voice tremor severity. *J Speech Lang Hear Res*. 2019;62:3689–3705. https://doi.org/10.1044/2019_JSLHR-S-19-0024.
23. Järveläinen H. Perception-based control of vibrato parameters in string instrument synthesis. *Proceedings of the International Computer Music Conference,* September 16–21, 2002. Goethenburg, Sweden. 287–294.
24. Zwicker E, Flottorp G, Stevens SS. Critical bandwidth in loudness summation. *J Acoust Soc Am*. 1957;29:548. https://doi.org/10.1121/1.1908963.
25. Moore B. Frequency difference limens for short duration tones. *J Acoust Soc Am*. 1973;54:237–246. https://doi.org/10.1121/1.1913640.
26. Moore BCJ. The role of temporal fine structure processing in pitch perception, masking, and speech perception for normal-hearing and hearing-impaired people. *J Assoc Res Otolaryngol*. 2008;9:399–406. https://doi.org/10.1007/s10162-008-0143-x.
27. Friberg A, Sundberg J. Just noticeable difference in duration, pitch and sound level in a musical context. *Proceedings of the 3rd International Conference for Music Perception and Cognition*. July 23–27, 1994. Liege, Belgium. 339–340.
28. Wier CC, Jesteadt W, Green DM. Frequency discrimination as a function of frequency and sensation level. *J Acoust Soc Am*. 1977;61:178–184. https://doi.org/10.1121/1.381251.
29. Barkmeier-Kraemer J, Story B. Conceptual and clinical updates on vocal tremor. *ASHA Lead*. 2010;15:16–19. https://doi.org/10.1044/leader.FTR2.15142010.16.
30. Lederle A, Barkmeier-Kraimer J, Finnegan E. Perception of vocal tremor during sustained phonation compared with sentence context. *J Voice*. 2012;26:668.e1–668.e9. https://doi.org/10.1016/j.jvoice.2011.11.001.
31. Daffern H, Brereton J, Howard D. The impact of vibrato usage on the perception of pitch in early music compared to grand opera. *Société Française d'Acoustique, Acoustics 2012*. April 2012; Nantes, France. ⟨http://hal.science/hal-00811314/document⟩.
32. Schonle J, Horan KE. The pitch of vibrato tones. *J Acoust Soc Amer*. 1980;67:246. https://doi.org/10.1121/1.383733.
33. d'Alessandro C, Castellengo M. The pitch of short-duration vibrato tones. *J Acoust Soc Amer*. 1994;95:1617–1630. https://doi.org/10.1121/1.408548.
34. Erickson ML. Can inexperienced listeners hear who is flat? The role of timbre and vibrato. *J Voice*. 2015;30:638.e9–638.e20. https://doi.org/10.1016/j.jvoice.2015.07.014.
35. Erickson ML, Gaskill CS. Can listeners hear how many singers are singing? The effect of listener's experience, vibrato, onset, and formant frequency on the perception of number of simultaneous singers. *J Voice*. 2012;26:817.e1–817.e13. https://doi.org/10.1016/j.jvoice.2012.04.011.
36. Horii Y, Hata K. A note on phase relationships between frequency and amplitude modulations in vocal vibrato. *Folia Phoniatr Logop*. 1988;40:303–311. https://doi.org/10.1159/000265924.
37. Horii Y. Acoustic analysis of vocal vibrato: a theoretical interpretation of data. *J Voice*. 1989;3:36–43. https://doi.org/10.1016/S0892-1997(89)80120-1.
38. Hajda JM. *The effect of time-variant acoustical properties on orchestral instrument timbres* [doctoral dissertation]. Los Angeles, CA: University of California; 1999.
39. Almeida A, Schubert E, Wolfe J. Timbre vibrato perception and description. *Music Percept*. 2021;38:282–292. https://doi.org/10.1525/mp.2021.38.3.282.
40. Nix J. Listener preferences for vibrato rate and extent in synthesized vocal samples. *Proc Meet Acoust*. 2015;23:035003. https://doi.org/10.1121/2.0000067.
41. Wooding R, Nix J. Perception of non-vibrato sung tones: a pilot study. *J Voice*. 2016;30:762.e15–762.e21. https://doi.org/10.1016/j.jvoice.2015.10.005.
42. Vatti M, Santurette S, Pontoppidan NH, et al. Maximum acceptable vibrato excursion as a function of vibrato rate in musicians and non-musicians. *Proc Acoust Soc Amer*. 2014;20:050002. https://doi.org/10.1121/1.4894272.
43. Reddy AA, Subramanian U. Singers' and nonsingers' perception of vocal vibrato. *J Voice*. 2015;29:603–610. https://doi.org/10.1016/j.jvoice.2014.09.022.
44. Howes P, Callaghan J, Davis P, et al. The relationship between measured vibrato characteristics and perception in Western operatic singing. *J Voice*. 2004;18:216–230. https://doi.org/10.1016/j.jvoice.2003.09.003.
45. Kreiman J, Antoñanzas-Barroso N, Gerrat BR. Integrated software for analysis and synthesis of voice quality. *Behav Res Methods*. 2010;42:1030–1041. https://doi.org/10.3758/BRM.42.4.1030.
46. Tanaka JW, Curran T, Sheinberg DL. The training and transfer of real-world perceptual expertise. *Psychol Sci*. 2005;16:145–151. https://doi.org/10.1111/j.0956-7976.2005.00795.x.
47. Granqvist S. The visual sort and rate method for perceptual evaluation in listening tests. *Logo Phoniatr Vocol*. 2003;28:109–116. https://doi.org/10.1080/14015430310015255.
48. Boersma P, Weenink D. Praat: doing phonetics by computer. (Version 6.3.08). Accessed February 10, 2023. Available at: ⟨http://www.praat.org/⟩.
49. Bates D, Mächler M, Bolker B, et al. Fitting linear mixed-effects models using lme4. *J Stat Softw*. 2015;67:1–48. https://doi.org/10.18637/jss.v067.i01.
50. Garner M, Lemon J, Fellows I, et al. Various coefficients of interrater reliability and agreement. (Version 0.84.1). Published January 26, 2019. Accessed October 13, 2022. Available at: ⟨https://cran.r-project.org/web/packages/irr⟩.
51. Powell MJD. *The BOBYQA Algorithm for Bound Constrained Optimization Without Derivatives*. Cambridge, England: Department of Applied Mathematics and Theoretical Physics, 2009.
52. American Speech-Language-Hearing Association. Evidence-based practice in communication disorders. 2005. Available at: ⟨http://www.asha.org/policy⟩. Accessed February 20, 2023.
53. Ragan K. Defining evidence-based voice pedagogy: a new framework. *J Sing*. 2018;75:157–160.
54. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiro Med*. 2016;15:155–163. https://doi.org/10.1016/j.jcm.2016.02.012.
55. Monson BB, Hunter EJ, Lotto AJ, Story BH. The perceptual significance of high-frequency energy in the human voice. *Front Psychol*. 2016;5:1–11. https://doi.org/10.3389/fpsyg.2014.00587.
56. Nandamudi S, Scherer RC. Airflow vibrato: dependence on pitch and loudness. *J Voice*. 2019;33:815–830. https://doi.org/10.1016/j.jvoice.2018.05.007.
57. Meyer D, Nix J, Okerlund D. Practical science in the studio, part 3: 'High-tech' strategies. *J Sing*. 2021;77:633–643.