Acoustic Parameters Critical for an Appropriate Vibrato

*Supraja Anand, *Judith M. Wingate, †Brenda Smith, and ‡Rahul Shrivastav, *†*Gainesville, Florida,* ‡*East Lansing, Michigan*

Summary: Background. A plethora of investigations have studied the acoustic characteristics of vibrato such as the rate, extent, onset (time from initiation of phonation until the first peak of vibrato), and periodicity. Despite extensive research, the degree to which various parameters of vibrato contribute to its acceptability remain unclear.

Purpose. The present study sought to determine the psychoacoustical relationship of mean fundamental frequency (f_0), modulation frequency (f_{f0m}), modulation depth (d_{f0m}), and intensity to the appropriateness or inappropriateness of vibrato.

Method. Phonation samples of eight voice majors singing at low, middle, and high pitches were obtained. A high fidelity vocoder (STRAIGHT; Kawahara, 1997) was used to resynthesize these vowels with systematic manipulations of f_{f0m} and d_{f0m} of the f_0 contours resulting in a total of 600 stimuli (8 singers $\times$ 3 pitches $\times$ 5 f_{f0m} levels $\times$ 5 d_{f0m} levels). Nine listeners (four experts and five students) evaluated these stimuli for appropriateness of vibrato at two different presentation levels (70 and 90 dB sound pressure level).

Results. Statistical analyses of the perceptual data suggest that appropriateness of vibrato tends to increase with mean f_0 and decrease with $d_{f0m.}$ Appropriateness of vibrato is greatest for f_{f0m} value of 6 Hz, but decreases both above and below this frequency.

Conclusion. perceived appropriateness of vibrato results from an interaction of mean f_0, f_{f0m}, and d_{f0m} of the vowel waveform.

Key Words: Vibrato–Opera–Modulation frequency–Modulation depth–Synthesis–Vocoder.

INTRODUCTION

Vocal vibrato is a common characteristic of vocal quality in Western singing styles. Seashore[1] in 1932 stated that "A good vibrato is a pulsation of pitch, usually accompanied with synchronous pulsations of loudness and timbre, of such extent and rate as to give a pleasing flexibility, tenderness, and richness to the tone." It is accepted in singing literature that poor vibrato is an indication of poor technique and inferior sound quality.[2–4]

Research over the years has characterized vibrato as a rhythmic modulation of the voice fundamental frequency characterized by rate, extent, onset, and periodicity. Rate and extent vary considerably depending on the genre,[5] style,[6,7] singer, and the emotional content of the song.[8–10] In general, vibrato rate or modulation frequency is defined by the rate of modulation about the mean f_0 (henceforth, referred to as f_{f0m}). Mean vibrato rates reported in various studies range from 4 to 7 Hz.[10–20] Vibrato extent has been described by how far (peak-to-trough) modulations vary relative to mean $f_{0,}$ or amplitude during the vibrato cycle. The most commonly reported vibrato extent, from a number of studies, amounts to ±1 semitone (ST) (approximately 6% of mean f_0).[14,16–19,21–23] To be consistent with psychoacoustical literature, vibrato extent is referred to as "modulation depth (d_{f0m})" in the present study.

Much of the above-mentioned work has been done on acoustic characteristics of vibrato. However, this examination provides limited information about vibrato because it does not describe the perceptual significance of these acoustic parameters. Researchers in vocal pedagogy, and the singing community generally agree on the notion that opera singers develop vibrato naturally during the vocal training period.[24,25] However, understanding the acoustic-perceptual relationship may gradually become more important as students and teachers start to access computer software programs that enable them to observe and measure their vibrato features from spectrographs or other displays.[18]

There have been very few investigations of how acoustic features such as f_{f0m} and d_{f0m} relate to perceived vibrato. Ekholm et al[13] examined vibrato from performances of singers with various voice types (four countertenors, seven tenors, and five baritones). The musical excerpt used in this study was from Mozart's concert aria, *Ch 'io mi scordi di te*. Vowel segments /a/, /i/, and /o/ of approximately 1–3 seconds in length were extracted from the 2-minute singing sample. These vowel phonations from singers (48 in total) were rated by panels of expert voice teachers, according to four perceptual criteria: "resonance/ring," "color/warmth," "clarity/focus," and "appropriate vibrato." For each voice sample and perceptual criteria, the judges assigned a score from 1 (poor) to 7 (excellent). Subjective ratings were related to objective measurements taken from acoustic analysis of the voice signal. The f_{f0m} of the vibrato samples ranged between 5.1 and 6.8 Hz, with a mean and standard deviation (SD) of 5.70 ± 0.45 and d_{f0m} ranged between 3.5% and 8% of mean $f_{0,}$ with a mean and SD of

Accepted for publication June 1, 2012.
Presented at the 39th Annual Symposium of the Voice Foundation: Care of the Professional Voice.
From the *Department of Speech, Language, and Hearing Sciences, University of Florida, Gainesville, Florida; †Department of Music, University of Florida, Gainesville, Florida; and the ‡Department of Communicative Sciences and Disorders, Michigan State University, East Lansing, Michigan.
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. 820.e19-820.e25
0892-1997/$36.00

http://dx.doi.org/10.1016/j.jvoice.2012.06.004

5.05% ± 1.38. Tenors received an average rating of 4.3 and countertenors received an average rating of 3.6 on appropriateness of vibrato. Perceptual data for baritones was discussed only for two singers (average ratings for appropriate vibrato, singer 1: 5.7; singer 2: 4.8). The authors concluded that the individual values of the average f_{f0m} and d_{f0m} of vibrato did not appear to have a significant influence on "appropriate vibrato" ratings. They also speculated that a combination of lower f_{f0m} and larger d_{f0m} would lead to lower ratings of appropriateness. Although this speculation seems reasonable, evaluation of their results is difficult because of lack of information and absence of statistical testing.

Howes et al[10] explored the acoustic-perceptual relationship of vibrato in Western Operatic singing using recordings of performances by internationally famous opera singers. Extracts chosen for this study included arias in Donizetti's *Lucia de Lammermoor*, and an emotional cadenza from a soprano aria, taken from Verdi's *Un ballo in maschera*. Measurements of f_{f0m} and d_{f0m} from different cadenzas were calculated manually from spectrograms created using *Matlab* (The Mathworks, Natick, MA). Expert listeners used a 10-cm Visual Analog Scale to indicate if f_{f0m} was "too fast/too slow" and "too wide/too narrow." Interestingly, the authors reported that f_{f0m} across all singers were similar (6.28–7.59 Hz). The d_{f0m} ranged from 0.41 to 1.58 ST. Judgments of f_{f0m} in the "too slow" (<3.9 cm) and "too fast" (>6.1 cm) category showed weak to moderate correlation ($r = 0.28$ and 0.37, respectively) with rate. Judgments of d_{f0m} in the "too wide" (<3.9 cm) and "too narrow" (>6.1 cm) ranges showed a moderate correlation ($r = 0.47$ and 0.30, respectively) with extent.

In summary, although published studies of vibrato have focused on f_{f0m} and d_{f0m}, there are several methodological differences such as sample size, differences in the selection of stimuli (live vs recorded stimuli, cadenzas used), stimuli that do not represent the entire range of vibrato rate and extent to be perceived as appropriate/inappropriate, differences in design, and the broad range of terminologies used to obtain perceptual data (eg, binary classifications such as good/bad, agreeable/nonagreeable, pleasant/unpleasant). These sources of variation make it difficult to draw valid conclusions regarding vibrato characteristics. Therefore, to identify such acoustic-perceptual relationships in a systematic manner, the present study investigated the psychoacoustical functions that relate mean f_0, f_{f0m}, d_{f0m}, and intensity to the appropriateness of vibrato. It may be useful to evaluate presentation level because perception of vibrato characteristics may be dependent on intensity in a subtle manner. This information can be beneficial in several ways, such as 1) audio-recording engineers can fine-tune a singer's vibrato for a listener and 2) a singer could adjust his/her environment (eg, microphone-mouth distance; placement of speakers) to produce desirable effects.

METHODS

Singers

Eight voice majors from four voice classifications namely bass, tenor, alto, and soprano sang the vowel /a/ at comfortable low, middle, and high pitches. The participants were allowed to choose a reference for these pitches using a tone played on the piano (tuned to A 440). For each pitch, five repetitions were obtained, each lasting approximately 5 seconds. The musical notes and the mean f_0 (Hz) measured through STRAIGHT algorithm are shown in Table 1. Sound recordings were made in a single-walled sound booth with a cardioid dynamic microphone (Shure SM 48) mounted on a stand. A mouth-microphone distance of 10 cm was maintained throughout the experimental trials. These signals were recorded using a portable Marantz PMD671 recorder at 44.1 kHz (16 bits per sample).

Stimuli

Speech synthesis allows an experimenter to manipulate acoustic variables in speech stimuli and assess their perceptual importance by conducting listening tests. A 1-second vowel segment was extracted from the temporal midpoint of the third trial for systematic manipulation of f_{f0m} and d_{f0m}. A custom-designed algorithm in *MATLAB* (version 7.0; The Mathworks, Natick, MA) was used to systematically manipulate f_{f0m} from 2 to 10 Hz in steps of 2 Hz and d_{f0m} from 2 to 26 Hz in steps of 6 Hz for each mean f_0. This was done by first extracting the f_0 contour using a high-precision algorithm (STRAIGHT;

TABLE 1.
Singers' Frequency

	Low f_0		Middle f_0		High f_0	
Voice Class/Singer	in Hz	Musical Note	in Hz	Musical Note	in Hz	Musical Note
Bass: singer 1	132	C_3	265	C_4	339	F_4
Bass: singer 2	106	G_2#	210	G_3#	289	D_4
Tenor: singer 1	131	C_3	194	G_3	298	D_4
Tenor: singer 2	172	F_3	266	C_4	359	F_4#
Alto: singer 1	215	A_3	327	E_4	662	E_5
Alto: singer 2	148	D_3	382	G_4	627	D_5#
Soprano: singer 1	216	A_3	433	A_4	875	A_5
Soprano: singer 2	273	C_4#	470	A_4#	736	F_5#

Notes: The musical note represents the reference from a piano tuned to A 440 and f_0 (Hz) represents the mean fundamental frequency extracted using an f_0 extraction algorithm in STRAIGHT.

Kawahara, 1997), then substituting it with another f_0 contour that had the desired modulation characteristics. The vowel was resynthesized with the modified f_0 contour, using a high-quality vocoder (STRAIGHT; Kawahara, 1997).[26] All other parameters, such as the formant frequencies, were retained from the original voice stimuli. The resulting vowel stimuli were highly natural and could not be discerned as being synthetically generated. The range for f_{f0m} and d_{f0m} manipulations were chosen to create a continuum of stimuli, which ranged from low amplitude/low frequency to high amplitude/high frequency of vibrato, in accordance with the values reported in the literature.[10–20] Thus, a total of 600 stimuli (8 singers $\times$ 3 pitches $\times$ 5 f_{f0m} $\times$ 5 d_{f0m}) were created. All the stimuli were resampled to 24 414 Hz to match the hardware requirements of the equipment used for the perceptual test. Figure 1 depicts an example of f_0 contour and variations in f_{f0m} and d_{f0m} for a stimulus with a mean f_0 of 470 Hz.

Because 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.[27] These ST values were used for all the subsequent statistical analyses.

$$\mathrm{ST} = 39.86 \times \log_{10}\left(\frac{f_0}{f_{0\mathrm{ref}}}\right) \quad (1)$$

For these calculations, the reference frequency ($f_{0\mathrm{ref}}$) was the lowest f_0 among all the stimuli (106 Hz).

Listeners

Five graduate students from Department of Speech, Language, and Hearing Sciences served as naïve listeners. Although these students had a background in music through participation in choirs, bands, and/or had instrumental lessons, they had less than 3 years of experience in Western classical music. Also, four expert judges (two speech-language pathologists with >30 years of experience in Western classical music and two professors from Department of Music) participated in the study. The age range of naïve listeners was 20–30 years and the age range of expert listeners was 50–70 years. All listeners were native speakers of American English and passed a hearing screening prior to the listening task.[28]

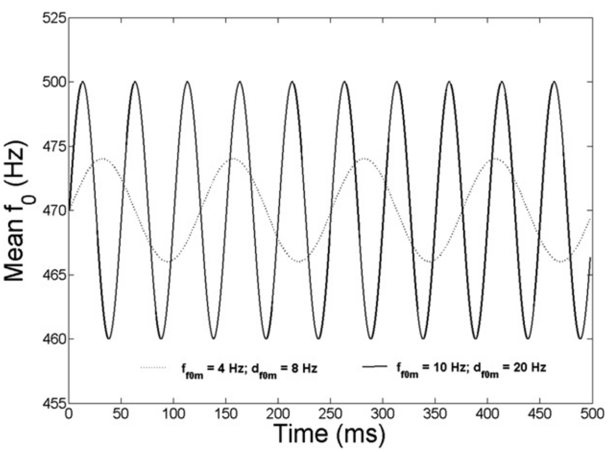

FIGURE 1. A graphical representation of f_{f0m} and d_{f0m} variations for a stimulus with mean f_0 of 470 Hz.

Instrumentation

Data acquisition procedures were controlled using the software *Sykofizx* through 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) at 70 and 90 dB sound pressure level or SPL in the right ear to avoid potential binaural interaction effects. The ER2 ear inserts are designed to deliver a flat frequency response between 100 Hz and 10 kHz at the eardrum.

Procedure

All the procedures used in this study were approved by the institutional review board at the University of Florida. Listeners responded "yes" or "no" using a computer interface to indicate whether the vibrato in the given stimuli was "appropriate" for current day opera singers. Listeners were reminded periodically to listen only to the vibrato, and not the overall quality or any other perceptual dimension in the stimuli. Each stimulus was presented at both the intensity levels in random order resulting in 1200 test stimuli and was repeated for three times (1200 $\times$ 3). Multiple responses were obtained for each stimulus for the reason that averaging across responses minimizes several errors and biases that can affect the data and the interlistener variability.[29] Thus, a total of 3600 stimuli were created. Given the large number of stimuli, listeners were required to take several short breaks during the listening task to avoid random errors caused by fatigue. Testing took approximately 6 hours for each listener, and was completed in three sessions of 2 hours each. Listeners were paid for participating in the experiment.

Statistical analyses

SPSS software (version 18.0; SPSS Inc., Chicago, IL) was used for all the analyses. A score of 1 was given for an "appropriate" rating and a score of 0 was given for an "inappropriate" rating. Although the assignment of 0 and 1 used to code "inappropriate" versus "appropriate" vibrato may seem categorical in nature, each stimulus was presented three times, and the stimuli were judged by nine listeners, resulting in a total of 27 independent responses for each stimulus. These responses were then averaged to result in a score that could vary between 0 and 1, reflecting the probability of a particular response (ie, appropriate or inappropriate). The large sample size (3600 stimuli) and normal distribution of data (confirmed *via* histogram analysis) suggested that parametric tests such as analysis of variance (ANOVA) were suitable to draw inferences from these data.

Inter- and intrajudge reliabilities were calculated for vibrato judgments using Pearson's product-moment correlation coefficient (r) and intraclass correlation coefficient (ICC). An independent sample t test was performed to compare the perceived "appropriate" scores of vibrato between the two listener groups (naïve and experts). A four-way ANOVA with post hoc comparisons using Bonferroni's correction was used to determine the effects of mean f_0, f_{f0m}, d_{f0m}, and stimulus intensity on perception of "appropriate" vibrato. Mean f_0, f_{f0m}, d_{f0m}, and stimulus intensity were the independent variables in the present study. The average vibrato judgment obtained from all listeners served as the dependent variable.

RESULTS

Reliability analysis

Interlistener reliability was computed using Pearson's correlation coefficient (r). "Interlistener reliability" was defined as the degree of consistency for the average ratings of "appropriate" vibrato between listeners for all the stimuli. The mean interlistener correlation coefficient was 0.69 (SD: 0.12). "Intralistener reliability" was defined as the degree of consistency within listeners between the three trials of vibrato rating for all the stimuli and was computed using ICC. The mean intrajudge correlation coefficient was 0.66 (SD: 0.15). Thus, the moderate-high correlations revealed that listeners made consistent judgments of "appropriate" vibrato.

Group differences

Table 2 depicts the mean vibrato judgments for both listener groups (naïve and experts). An independent sample t test revealed no significant difference ($t = -1.874$, $P = 0.061$) between the two listener groups in mean ratings of "appropriate" vibrato. Because there was no significant difference between naïve and expert listeners, mean vibrato judgments groups were combined for ANOVA testing.

Main effects

Mean f_0. Figure 2 demonstrates the significant main effect found for mean f_0 ($F(23,3576) = 48.37$, $P < 0.001$). When averaged across all f_{f0m} and d_{f0m}, voices with a high f_0 (alto and soprano singers) were perceived to have more "appropriate" ratings of vibrato when compared with those with a low f_0 (bass and tenor singers).

f_{f0m}. Significant main effects were also found for f_{f0m} ($F(4,3580) = 481.43$, $P < 0.001$), voice class ($F(3,3580) = 116.91$, $P < 0.001$), and the interaction between f_{f0m} and voice class ($F(12,3580) = 3.80$, $P < 0.001$) as illustrated in Figure 3. Modulation frequency (f_{f0m}) of 6 Hz was perceived to be "more appropriate" when compared with the other modulation frequencies. Voice classes' bass and tenor received similar ratings, but these differed from those for alto and soprano. At low f_{f0m} of 2 Hz, all the voice classes received lower ratings for "appropriate" vibrato.

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

Listener Type	N	Mean Vibrato Judgments	SD
Naïve	3600	0.3824	0.31999
Expert	3600	0.3971	0.34582

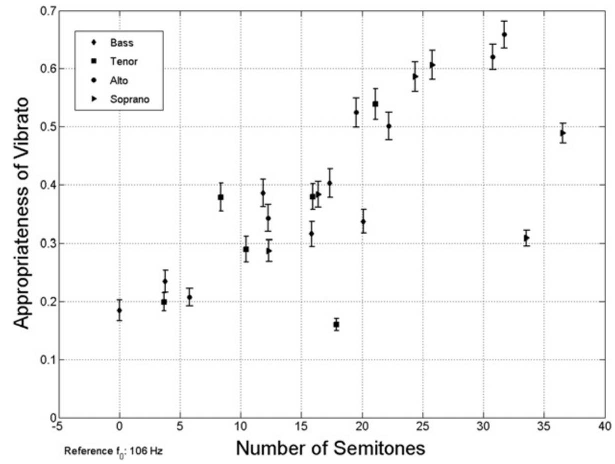

FIGURE 2. Vibrato judgments plotted against the mean f_0 in semitone scale. *Note:* Symbols indicate the mean probability of detection (or percentage of "appropriate" responses) averaged across listeners and bars indicate the standard error of the mean (SEM).

d_{f0m}. As with f_{f0m}, significant main effects were found for d_{f0m} ($F(4,3580) = 93.51$, $P < 0.001$), voice class ($F(3,3580) = 84.59$, $P < 0.001$), and the interaction between d_{f0m} and voice class ($F(12,3580) = 5.20$, $P < 0.001$) as illustrated in Figure 4. Modulation depths of 2 and 8 Hz (small d_{f0m}s) were perceived as more "appropriate." The perceived appropriateness of vibrato tends to decrease when d_{f0m} exceeds 8 Hz. However, this decline is greatest for the bass and tenor voices, and somewhat less for alto and soprano voices.

Intensity. As shown in Figure 5, intensity levels did not reach significance ($F(1,3592) = 0.225$, $P = 0.635$) on the perceived appropriateness of vibrato. However, there was a significant effect of voice class ($F(3,3592) = 75.73$, $P < 0.001$). Voice classes' alto and soprano were perceived to be more "appropriate"

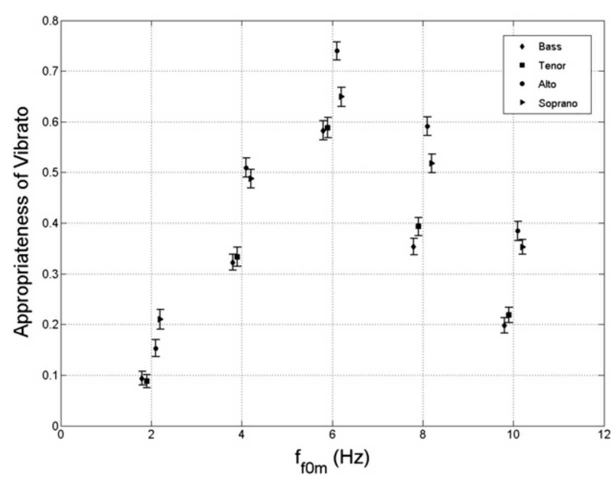

FIGURE 3. Effect of f_{f0m} on appropriateness of vibrato. *Note:* Symbols indicate the perceived judgments averaged across listeners and bars indicate SEM. For the purposes of clarity, the data points for different voice classes are shifted by ±0.1 offset around the actual f_{f0m} values.

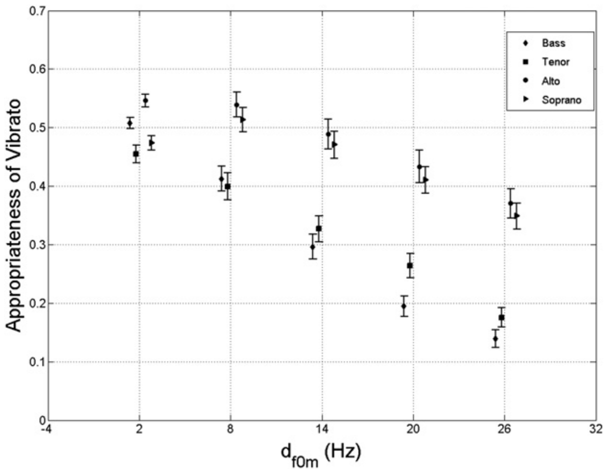

FIGURE 4. Effect of d_{f0m} on appropriateness of vibrato. *Note:* Symbols indicate the perceived judgments averaged across listeners and bars indicate SEM. For the purposes of clarity, the data points for different voice classes are shifted by ±0.1 offset around the actual d_{f0m} values.

than bass and tenor at both the intensity levels. The interaction between intensity and voice class did not reach significance ($F(3,3592) = 1.197$, $P = 0.309$).

The interactions as discussed under the Main Effects section describe interaction of each of the independent variable with the voice class. Further analyses were made to explore the relationship between the three significant independent variables (f_0, f_{f0m}, and d_{f0m}).

Interaction effects

Because there was a significant main effect of mean f_0, the vibrato judgments were normalized for f_0 as follows:

$$\widehat{a}_i = \frac{a_i}{\max(\overline{a})} \quad (2)$$

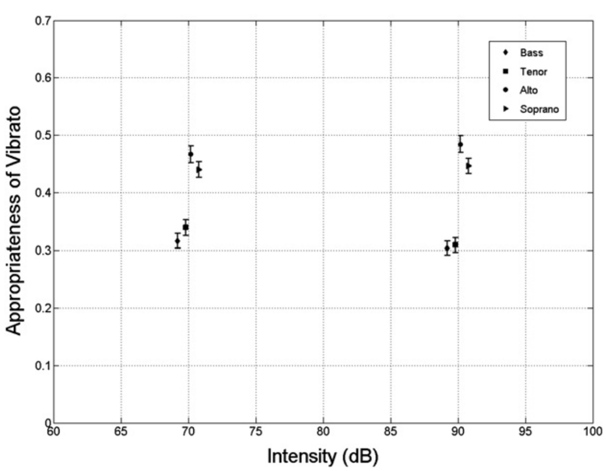

FIGURE 5. Effect of intensity on appropriateness of vibrato. *Note:* Symbols indicate the perceived judgments averaged across listeners and bars indicate SEM. For the purposes of clarity, the data points for different voice classes are shifted by ±0.1 offset in the abscissa.

where $\widehat{a}_i$ is the normalized appropriateness score; a_i is the mean appropriateness score across all the listeners for that f_{f0m} and d_{f0m} of a single mean f_0; and $\max(\overline{a})$ is the maximum appropriateness score for all the stimuli with the same f_0 but with different f_{f0m} and d_{f0m}.

Interaction 1: Mean f_0 versus f_{f0m}. A significant interaction was observed between mean f_0 and f_{f0m} ($F(92,3480) = 8.102$, $P < 0.001$) as illustrated in Figure 6. It is evident that at all mean f_0, f_{f0m} of 6 Hz was perceived to have more "appropriate" vibrato. It also appears that as the mean f_0 increased, the range of f_{f0m} that was perceived to have "appropriate" vibrato also increased. For example, for mean f_0 less than 25 ST, only a narrow range around 6 Hz f_{f0m} was considered to be appropriate; however, for mean f_0 greater than 25 ST, f_{f0m} range between 4 and 8 Hz was also perceived to have "appropriate" vibrato.

Interaction 2: Mean f_0 versus d_{f0m}. There was also a significant interaction between mean f_0 and d_{f0m} ($F(92,3480) = 2.18$, $P < 0.001$) as depicted in Figure 7. Irrespective of the mean f_0, d_{f0m} of 2 and 8 Hz was perceived to have more "appropriate" vibrato. However, at high mean f_0, the range of d_{f0m} that was perceived to have "appropriate" vibrato also increased, that is, at high mean f_0, all the d_{f0m} received higher ratings of appropriateness.

Interaction 3: f_{f0m} versus d_{f0m}. Figure 8 reveals that a final significant interaction was found between f_{f0m} and d_{f0m} ($F(16,3575) = 30.76$, $P < 0.001$). It is evident from the figure that at low d_{f0m}s (2 Hz), all modulation frequencies are perceived to be equally appropriate. At higher d_{f0m}s, a clear preference is seen for 6 Hz f_{f0m}. Also, there is a steeper decline and a lower rating of appropriateness for lower f_{f0m} (<6 Hz) when compared with higher f_{f0m} (8 and 10 Hz). Similarly, for a 2 Hz f_{f0m}, there is no effect of d_{f0m}, whereas at higher f_{f0m}'s such as 8 and 10 Hz, there was an effect of d_{f0m}, that is, lower d_{f0m} was perceived to be more appropriate (<8 Hz).

DISCUSSION

A vast majority of research has been conducted either on acoustic characteristics of vibrato or on perceptual features, but very

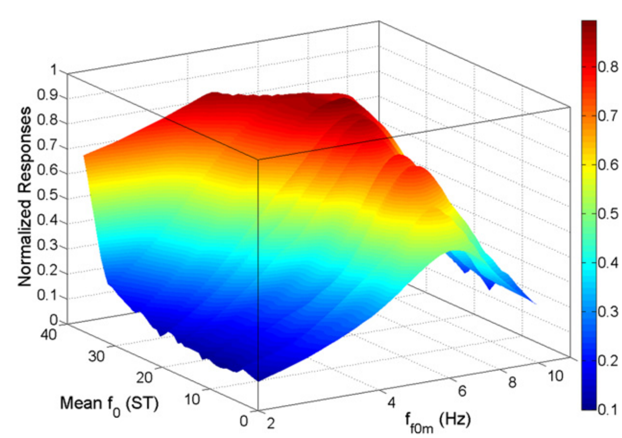

FIGURE 6. A three-dimensional representation of the interaction between mean f_0 and f_{f0m} on appropriateness of vibrato.

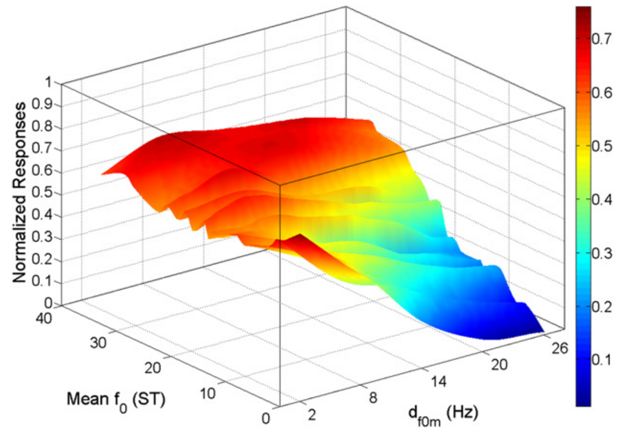

FIGURE 7. A three-dimensional representation of the interaction between mean f_0 and d_{f0m} on appropriateness of vibrato.

few studies have explored the link between these two aspects. The purpose of this study was to examine the association between acoustic measures of vibrato such as f_{f0m}, d_{f0m}, and their perceptual significance. Because the ideal vibrato parameters vary by musical era and genre, this study focused on listener's judgment based on current day opera singers. Perceptual labels of vibrato quality such as "good" or "bad," "pleasing" or "unpleasing" to the ear are very subjective, and can differ based on the individual preferences of the listener. Therefore, the label "appropriate" was chosen in this study. Vibrato parameters were studied on synthetic samples as speech synthesis allows systematic manipulation of independent variables of interest. In addition, vibrato was studied on vowel phonations of /a/ given that singing/voice teachers generally use this vowel to teach vibrato if there was a problem (eg, excessive vibrato).

Moderate reliability in the listener judgments could have resulted because judging vibrato quality while overlooking other voice properties, such as vocal stability, intensity, or timbre might be difficult. There was no effect of listener experience (naïve vs expert) in this study. To date, there has been no study that has compared listener experience on perception of vibrato (Experiments have almost always used expert listeners for making perceptual judgments). It should be noted that vibrato is typically used for expressive purposes by singers, and judges would be expected to relate to it as such. The judges in the present study, however, were required to evaluate vibrato in isolated phonation, and not within a musical phrase. Thus, their experience in assessing vibrato may not have been exploited to its fullest capability.

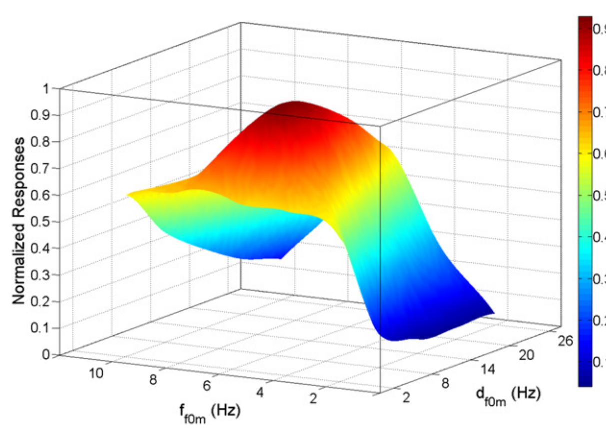

FIGURE 8. A three-dimensional representation of the interaction between f_{f0m} and d_{f0m} on appropriateness of vibrato.

Examination of mean f_0 results revealed that as f_0 increased, perception of vibrato was considered more "appropriate." One possible explanation of this finding could be attributed to singer effects. Vibrato from alto and soprano singers received higher ratings of appropriateness when compared with bass and tenor singers. Irrespective of voice class, a 6 Hz f_{f0m} vibrato was perceived to be the "most appropriate." This finding is consistent with previous literature on vibrato rate.[10–20] Lower extents d_{f0m} of 2 and 8 Hz are preferred by the judges (naïve and expert) in the present study, which mirrors the changing preferences of the general population toward Western classical music.[10,30] Intensity of stimulus presentation did not have an effect on the perceived appropriateness. It is important to note that these results are based on synthetic samples that were specifically designed to manipulate f_{f0m} and d_{f0m} while keeping other variables constant. However, when a singer makes volitional changes, additional voicing and vibrato characteristics may change, and these could further alter listener judgments of appropriateness.

Based on the interaction between mean f_0 and f_{f0m}/d_{f0m}, it was observed that at higher mean f_0, a broader range of vibrato rates and extents were considered "appropriate." In contrast, at low mean f_0, only a very narrow range of modulation characteristics resulted in perception of appropriate vibrato. Such nonlinear effects are typical in many psychoacoustic phenomena, such as critical bands.[31,32] This suggests that the nonlinear characteristics are inherent in the auditory system, and the perception of vibrato could be explained using the same phenomena.

The interaction effect found between f_{f0m} and d_{f0m} revealed that listeners rated the vibrato samples as inappropriate at low values of f_{f0m} and/or d_{f0m}. This result suggests that an f_{f0m} and/or d_{f0m} of 2 Hz falls below the detection threshold, and hence listeners did not perceive the vibrato to rate it as appropriate. Also, lower extents are preferred, but not lower rates. A vibrato rate of lower than 4 Hz is commonly referred to as "wobble" and is not preferred by singing pedagogues. Accordingly, the boundaries of rate and extent scale are encouraging because these segments of the scale are of most interest from a singing education perspective. Although Ekholm et al's data revealed that average values of f_{f0m} and d_{f0m} of vibrato did not appear to have a significant influence on "appropriate" vibrato ratings, findings from the present study show that f_{f0m} and d_{f0m} are important. Ekholm et al's study also speculated that a combination of lower f_{f0m} and larger d_{f0m} would lead to lower ratings of appropriateness. This finding is consistent with the present study.

The major perceptual effect of the vibrato is dependent on the f_{f0m}. However, f_0 modulations can lead to proportional variation of harmonics, thereby resulting in amplitude modulations that may be relevant to the perceived quality of vocal vibrato. The present study did not directly assess the amplitude modulations.

Hence, further research needs to be done to elucidate the exact mechanisms of amplitude vibrato.

CONCLUSION

Findings of vibrato characteristics from this study might have applications in the singing studio in terms of providing feedback to a student. Relating perceptual judgments to acoustic measurement has the potential to provide a basis for a more systematic and consistent pedagogical approach. Future research should explore the effect of emotions in assessing vibrato and using longer samples for listening experiments. Commercial audio engineers in the music industry could thus make use of computer software to enhance vibrato quality of a specific sound sample based on perceptual data collected from the expert judges.

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. Seashore CE, ed. *The Vibrato. University of Iowa Studies in the Psychology of Music*, 1932;Vol I. Iowa City, IA: University of Iowa; 1932.
2. Vennard W. *Singing, the Mechanism and the Technique*. 2nd ed. New York, NY: Carl Fischer; 1967. 192–208.
3. Miller R. *The Structure of Singing*. New York, NY: Schirmer; 1996.
4. Sundberg J. Acoustic and psychoacoustic aspects of vocal vibrato. In: Dejonckere P, Hirano M, Sundberg J, eds. *Vibrato*. San Diego, CA: Singular Publishing Company; 1995:35–62.
5. Carter NJ, Hopkin JA, Dromey C. Volitional control of vibrato. *J Singing*. 2010;67:9–17.
6. Easley E. A comparison of the vibrato in concert and opera singing. In: Seashore CE, ed. *The Vibrato*. Iowa, IA: University of Iowa Press; 1932: 269–275.
7. Hakes J, Shipp T, Doherty ET. Acoustic properties of straight tone, vibrato, trill and trillo. *J Voice*. 1987;1:148–156.
8. Gabrielsson A, Juslin PN. Emotional expression in music performance: between the performer's intention and the listener's experience. *Psychol Music*. 1996;24:68–91.
9. Sundberg J, Iwarsson J, Hagegard H. A singer's expression of emotions in sung performance. In: Fujimura O, Hirano M, eds. *Vocal Fold Physiology: Voice Quality Control*. San Diego, CA: Singular Publishing Group, Inc.; 1995:217–231.
10. Howes P, Callaghan J, Davis P, Kenny D, Thorpe W. The relationship between measured vibrato characteristics and perception in Western operatic singing. *J Voice*. 2004;18:216–230.
11. Ramig LA, Shipp T. Comparative measures of vocal tremor and vocal vibrato. *J Voice*. 1987;1:162–167.
12. Prame E. Measurements of the vibrato rate of ten singers. *J Acoust Soc Am*. 1994;96:1979–1984.
13. Ekholm E, Papagiannis GC, Chagnon FP. Relating objective measurements to expert evaluation of voice quality in Western classical singing: critical perceptual parameters. *J Voice*. 1998;12:182–196.
14. Titze IR, Story B, Smith M, Long R. A reflex resonance model of vocal vibrato. *J Acoust Soc Am*. 2002;111:2272–2282.
15. Bretos J, Sundberg J. Measurements of vibrato parameters in long sustained crescendo notes as sung by ten sopranos. *J Voice*. 2003;17:343–352.
16. Dromey C, Carter N, Hopkin A. Vibrato rate adjustment. *J Voice*. 2003;17: 168–178.
17. Mendes AP, Rothman HB, Sapienza C, Brown WS. Effects of vocal training on the acoustic parameters of the singing voice. *J Voice*. 2003;17:529–543.
18. Mitchell HF, Kenny DT. The impact of 'open throat' technique on vibrato rate, extent and onset in classical singing. *Logoped Phoniatr Vocol*. 2004; 29:171–182.
19. Murbe D, Zahnert T, Kuhlisch E, Sundberg J. Effects of professional singing education on vocal vibrato—a longitudinal study. *J Voice*. 2007;21: 683–688.
20. Prame E. Vibrato extent and intonation in professional Western lyric singing. *J Acoust Soc Am*. 1997;102:616–621.
21. Horii Y. Acoustic analysis of vocal vibrato: a theoretical interpretation of data. *J Voice*. 1989;3:36–43.
22. Sundberg J. *The Science of the Singing Voice*. Dekalb, IL: Northern Illinois University Press; 1987.
23. Keidar A, Titze I, Timberlake C. Vibrato characteristics of tenors singing high C's. In: Laurence V, ed. *Transcripts of the 13th Symposium Care of the Professional Voice Part I*. New York, NY: The Voice Foundation; 1984:105–110.
24. Bjorklund A. Analyses of soprano voices. *J Acoust Soc Am*. 1961;33: 575–582.
25. Nair G. *Voice Tradition and Technology: A State-of-the-Art Studio*. San Diego, CA: Singular Publishing Group; 1999.
26. Kawahara H. Speech representation and transformation using adaptive interpolation of weighted spectrum: vocoder revisited. *Acoust Speech Signal Process*. 1997;2:1303–1306.
27. Baken RJ, Orlikoff RF. *Clinical Measurement of Speech and Voice*. San Diego, CA: Singular Publishing Group; 1993.
28. American National Standards Institute. *Methods for Manual Pure-Tone Threshold Audiometry*. ANSI S3.21–2004. New York, NY: American National Standards Institute; 2004.
29. 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.
30. Strempel EL. The shifting aesthetics of vibrato. *J Singing*. 2006;62: 405–411.
31. Zwicker E, Flottorp G, Stevens SS. Critical bandwidth in loudness summation. *J Acoust Soc Am*. 1957;29:548–557.
32. Zwicker E, Fastl H. *Psychoacoustics: Facts and Models*. New York, NY: Springer-Verlag; 1990.