

TIMBRE VIBRATO PERCEPTION AND DESCRIPTION

ANDRÉ ALMEIDA, EMERY SCHUBERT, & JOE WOLFE
The University of New South Wales, Sydney, Australia

IN MUSIC, VIBRATO CONSISTS OF CYCLIC VARIATIONS in pitch, loudness, or spectral envelope (hereafter, “timbre vibrato”—TV) or combinations of these. Here, stimuli with TV were compared with those having loudness vibrato (LV). In Experiment 1, participants chose from tones with different vibrato depth to match a reference vibrato tone. When matching to tones with the same vibrato type, 70% of the variance was explained by linear matching of depth. Less variance (40%) was explained when matching dissimilar vibrato types. Fluctuations in loudness were perceived as approximately the same depth as fluctuations in spectral envelope (i.e., about 1.3 times deeper than fluctuations in spectral centroid). In Experiment 2, participants matched a reference with test stimuli of varying depths and types. When the depths of the test and reference tones were similar, the same type was usually selected, over the range of vibrato depths. For very disparate depths, matches were made by type only about 50% of the time. The study revealed good, fairly linear sensitivity to vibrato depth regardless of vibrato type, but also some poorly understood findings between physical signal and perception of TV, suggesting that more research is needed in TV perception.

Received: February 6, 2020, accepted October 10, 2020.

Key words: vibrato, perception, timbre, loudness, depth

VIBRATO IS THE CYCLIC VARIATION OF acoustical parameters in a pitched sound, with the variation having typically a frequency of several hertz. It is an important decoration in musical expression on most wind and bowed string instruments and the voice (e.g., Geringer, MacLeod, Madsen, & Napoles, 2014; Papich & Rainbow, 1974; Zhang, Bocko, & Beauchamp, 2015). Seashore (1938) wrote, “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” (p. 33).

In vibrato, periodic loudness, pitch, and timbre variations often occur together. In musical instruments, for example, nonlinear effects in the bow-string interaction or the reed or lips of a player have the result that changes in the amplitude of a vibration are also expected to produce variations in spectral envelope (e.g., Benade, 1976). Clipping of a sine wave is a simple example (which shares features with a clarinet’s reed when it beats against the mouthpiece, or the effects pedal of an electric guitar).

The co-occurrence of vibrato types creates problems when musicians communicate about vibrato. For example, when one describes vibrato on a violin, it is commonly understood in terms of oscillations in pitch, but it is difficult to have these oscillations without correlated spectral and loudness changes because the spectral response of the instrument is a strong function of fundamental frequency (e.g., Gough, 2005). There has been little research investigating the perceptual relationship between different types of vibrato. Some experiments suggest that pitch fluctuations (hereafter “pitch vibrato”) have the largest influence on the vibrato percept (Horii & Hata, 1988, as cited in Desain, Honing, Aarts, & Timmers, 1999). Others (Hajda, 1999) concluded that removing variations in spectral ratio from a musical tone has bigger consequences than removing either pitch or amplitude variations. As Seashore (1937, Chapter 4) writes, “The vibrato is always heard as of very much smaller extent than it is in the physical tone.” For example, a vibrato with a measured amplitude of one semitone does not seem to the listener to be nearly so wide. Katok (2016, p. 16) stated that the perception of vibrato depends upon recognizing its presence as well as being able to differentiate between the different types of vibrato.

Absent in the literature is the systematic investigation of the nature of different types of vibrato and the extent of their relative perceptual sensitivities. If different types of vibrato could be systematically separated, would they be perceived differently along type and intensity parameters?

The current study compares perceptions of cyclic variations in loudness and spectral amplitude, hereafter called loudness vibrato and timbre vibrato types, respectively. Loudness vibrato is produced here by varying the amplitude of all harmonics in a periodic tone proportionally. Timbre vibrato is produced here by a cyclic change in spectral slope, with the overall amplitudes

being then adjusted to keep the loudness constant. Pitch vibrato was not included, because of the need to limit the duration of experimental sessions and because pitch vibrato has been more extensively studied.

We investigate timbre vibrato perception through three questions over three experiments. Two of these experiments are reported here. They investigated the perception of manipulated vibrato test tones in terms of depth of vibrato (sometimes referred to as “extent,” Prame, 1997) and type of vibrato (loudness and timbre)—in comparison to a reference tone—using an experimental paradigm of selecting the best matching, most similar option from a selection of tones to a reference tone. Timbre and related words were never mentioned, and subjects were not advised what “similar” or “match” might mean.

The first experiment manipulated vibrato depth of test tones each with the same vibrato type (e.g., both loudness vibrato and no timbre vibrato). We wanted to see which of the test tones participants would choose that best match a reference tone with a different vibrato type (and so in the example, timbre vibrato), and the level of precision with which the matching could be performed. The second experiment held depth of the test tones constant, and forced participants to choose which vibrato type best matched the reference tone, even if the reference tone has a considerably different vibrato depth to the test tones.

The experiments took about 25 minutes to complete and used a graphical interface on a desktop computer playing synthesized tones through closed headphones. Planning to keep the sessions to this time limit was considered important in maintaining concentration and aiding recruitment. However, it did require limits to the number of iterations in sequential judgments.

Experiment 1. Vibrato Depth Matching

AIM

Experiment 1 aimed to quantify sensitivity to the depth of two types of vibrato: loudness and timbre using an adaptive paradigm, with stimuli selected based on sensitivity of previous participant response.

METHOD

Materials. Tones of constant duration and pitch were synthesized with different depths of loudness vibrato or timbre vibrato (depth being one of two independent variables), each type synthesized so as to minimize vibrato of the other type. Tones were designed as a sum of time-varying sinusoidal components with $N = 15$ harmonic partials. They were calculated explicitly in python+numpy and generated at a sample rate of

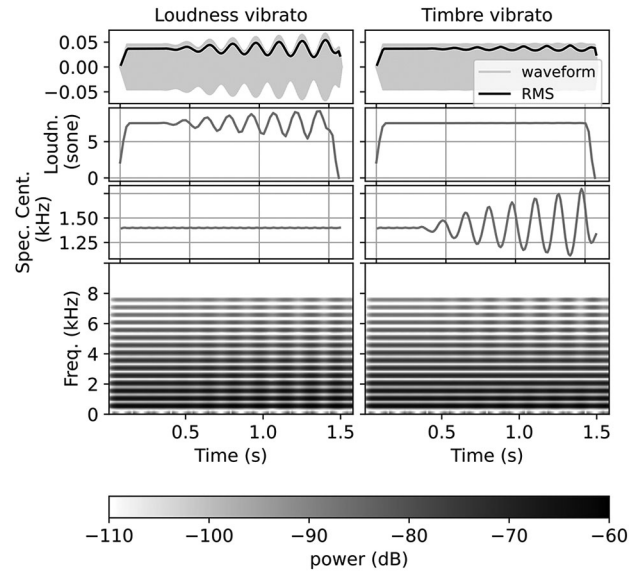


FIGURE 1. Two samples used in the perceptual test: Left panels, a loudness vibrato of depth 0.5, right panel, a timbre vibrato 0.5 in depth. From top to bottom, Waveform and RMS amplitude, loudness according to (PsySound3, window size of 93 ms = 4094 samples), spectral centroid and spectrogram.

44.1 kHz. Spectral centroid, SC (which can be visualized as the center of mass of the spectrum), is taken as the controlled variable for timbre because of its close association with perceived brightness (Almeida, Schubert, Smith, & Wolfe, 2017).

Examples are shown in Figure 1, and eight of the possible tones used in the experiments are provided as accompanying media. At any time, tones have a linear spectral slope (as measured in decibels/harmonic number). For loudness vibrato, the slope remains constant in time at -0.95 dB per harmonic, with the spectral centroid nearly constant at approximately 1350 Hz. For timbre vibrato, the slope oscillates about this value. The amplitude of each harmonic is multiplied by a function of time $f(t)$ that oscillates around the value 1, growing to an amplitude V_d . The amplitude of this oscillation is zero from 0 to 0.3 seconds, then grows linearly to a maximum at 1.5 seconds, then quickly falls to zero at 1.6 seconds (see Figure 1). The vibrato depth parameter V_d , the second of two independent variables, corresponds to the fractional modulation of loudness or spectral centroid. These definitions are somewhat arbitrary and there is no expectation that a same value of depth is perceptually similar for timbre or loudness vibrato. It is, rather, a way of quantifying these variations so that later we can relate the perceptual depths of the two parameters.

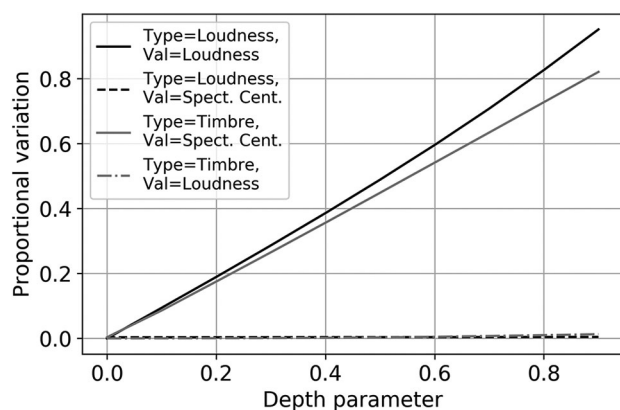


FIGURE 2. Maximum proportional variation of Loudness in sones and Spectral Centroid (“Val” in legend) attributed to each vibrato depth parameter (x -axis). Variation is displayed as a fraction of the average value. (Loudness is measured in sones according to the Moore model.)

In timbre vibrato, the spectral slope varies with time in such a way that the spectral centroid has the same function of time $f(t)$. The amplitude of all the harmonics is then multiplied by a single factor such that the loudness of the resulting sound according to the Moore and Glasberg (MG) model in Pysound 3 (Cabrera, 1999) was constant over time. (Because this model is based on averaging over varying population response, the timbre vibrato tones may have had a loudness vibrato component for some participants.)

The range of V_d was limited to 0 to 1 in both cases. In an informal, preliminary experiment, a small number of participants attested that these ranges had comparable perceptual magnitudes.

All the stimuli had a duration of 1.6 second with a fundamental frequency of 500 Hz. None of the vibrato tones included fluctuation of the pitch: the frequency of all harmonics remained stable throughout the tone. They had linear starting and finishing transients with durations of 50 and 20 ms, respectively. The variation in loudness and spectral centroid as functions of vibrato depth are shown in Figure 2.

A selection of tones used in the experiment can be found in the additional media for this article and at <http://newt.phys.unsw.edu.au/jw/vib1.html>.

PROCEDURE

The experiment was performed on a computer graphical interface that presented a reference tone followed by three test tones, as illustrated in Figure 3. The participant’s primary task was to select one of the three test tones that best matched (sounds “closest” to) the

reference tone. The depth matching experiment was conceived as an iterative, adaptive process (Leek, 2001). Participants were presented with a reference tone (Figure 3), either of timbre or loudness vibrato type, together with three test tones, all with the other vibrato type but each having a different depth. The reference depth parameter was designed to be a value chosen at random between 0 and 1 with uniform probability, although for a few subjects, when comparing vibrato of same type, the distribution was biased towards higher depths. One of the depths of the test tones (d_t) is chosen as the closest to the reference from the previous iteration. The other two tones have a depth d_t/s and d_t*s , where s is a parameter starting at 2, and which was reduced at each iteration by a factor of 1.3, if the participant reported a high confidence (“Certain,” “Fairly confident,” or “Slightly confident”) but not reduced for low (“Not confident” or “Guessing”). This process was iterated up to five times for any given reference tone and, again, the most dissimilar test tones in terms of depth. Due to an error in the program, the last interval range between test tones could, for a consistently confident participant, become larger in the fifth than the fourth iteration. The data from such diverging tones made little difference to the results, but the participants corresponding to the faulty cases were removed from the experiment. For each iteration, the chosen tone and the chosen depth were recorded along with the degree of confidence.

The tones were presented with a binaural, closed, around-the ear headset (Sennheiser HD280 pro), at levels of 46–49 dB (A). The sound level inside a headphone volume with no signal from the computer was 39 dB (A) measured with a Bruel and Kjaer model 2250-S sound level meter. Tests for harmonic distortion in the computer-headphone reproduction system using a pure sine wave from the computer found that any harmonic components due to distortion were at least 40 dB below the fundamental.

PARTICIPANTS

One hundred and seventy-eight students (107 female, 69 male) enrolled in a Music Psychology course participated in the study, approved by the Human Ethics Committee of UNSW Sydney. Forty-nine were music students, 84 others had music experience, and 43 had little music experience. One hundred and seventy-one of the participants were aged between 18 and 23; seven were older than 23. They earned course credit in return for completing the experiments. The topics timbre and brightness had not been discussed in their course at the

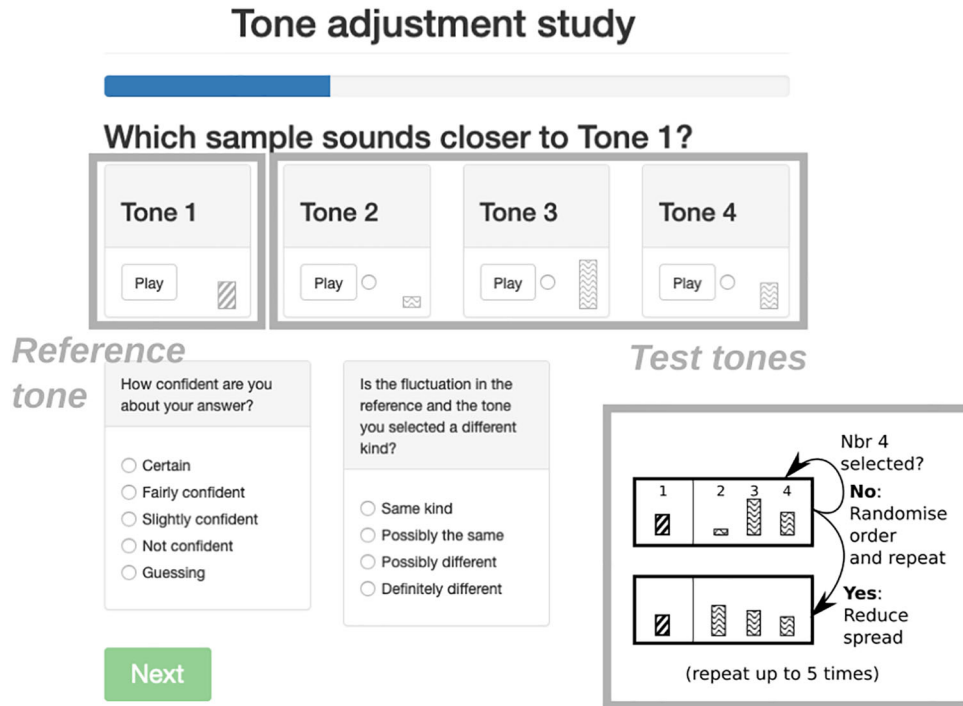


FIGURE 3. Screenshot of the interface for the depth matching experiment with annotations. The light grey colored text is added in this figure to indicate that tone 1 is the reference tone and tones 2, 3, and 4 are the test tones from which the participant has to make a selection. The small shaded grey rectangles (also not in the experiment) are visual representations of the possible types (fill pattern) and depth parameters (height) of the tones: In this experiment, test tones (tones 2, 3, and 4) are always of the same type, but the type of tone 1 can be the same or different from the test tones. The depth parameter of tone 1 is between the minimum and maximum depth of the test tones. Inset: Example of two successive trials of experiment 1: If the participant chooses the tone that is closer in depth parameter to the reference tone (1), then the spread of depths in the test tones (2-4) is reduced, otherwise the same depths are used but the order of the test tones is randomised.

time of the experiment. The experiment session included the perceptual test on the scaling of brightness (Almeida et al., 2017). All participants completed the session consisting of the sequence of experiments:

- Adjust loudness of stationary test tone to match reference tone with different spectrum. (5 trials, used in Almeida et al., 2017)
- Vibrato type similarity experiment (select vibrato tone closest to reference: 5 trials) EXP 2
- Adjust loudness of stationary test tone to twice the reference tone (5 trials, used in Almeida et al., 2017)
- Vibrato depth matching experiment (select 1 of 3 vibrato tones closest to reference: 2 trials, 5 iterations each) EXP 1
- Adjust brightness of stationary tone to twice the test tone (Almeida et al., 2017, 5 trials)
- Vibrato depth matching experiment (2 trials, 5 iterations each) EXP 1

- Adjust brightness of stationary tone to half the test tone (Almeida et al., 2017, 5 trials)
- Vibrato depth matching experiment (2 trials, 5 iterations each) EXP 1
- Free comments. (not reported here)

Results

In this experiment, each participant was led iteratively to judge the depth of one vibrato type that best matched a given depth of the same or of the other vibrato type reference tone. In Figures 4 and 5, both axes show the vibrato depth parameter, which approximately equals the proportional change in the loudness or, for timbre vibrato, the proportional change in spectral centroid. Figure 4 shows the results when the test and reference vibratos were of the same type. Consider any one of the dots on the graph. Here a participant heard a particular reference tone with a loudness depth vibrato shown on

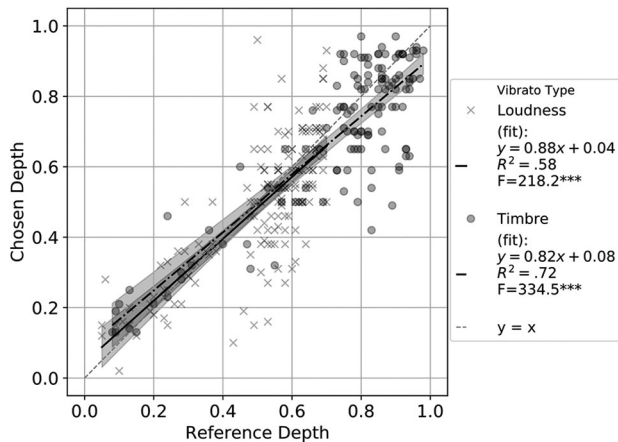


FIGURE 4. Depth parameter of the final match in the perceptual test plotted against the depth of the reference tone of the same vibrato type: Participants identify similar depths, but not with high precision and consistency. The diagonal dashed thin line corresponds to the idealized same depth in both reference and chosen tone. (For timbre, depth of 0.1 means that the maximum variation from the average value of spectral centroid is 10%; for loudness it means a peak change in loudness of 10%). Due to a programming error, the distribution of timbre reference depth was over-represented at higher values.

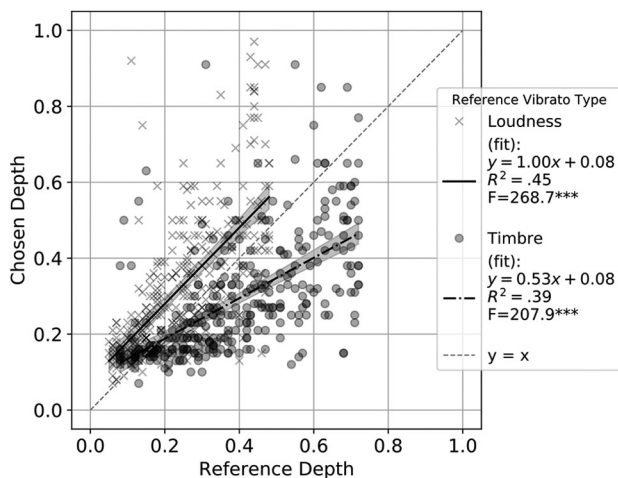


FIGURE 5. Vibrato depth of the final chosen tone plotted against the vibrato depth of the reference tone for test tones with different type of vibrato from the reference. For circles, the reference is timbre vibrato, for crosses, loudness. When the two vibratos are judged to have equal perceived depth, the depth parameter of timbre vibrato is larger than that for the loudness vibrato. Lines are linear fits to each of the reference tones, and light dashed line corresponds to equal depth in both tones.

the x -axis for that dot, and through the adaptive procedure, finally selected a tone with a loudness depth value indicated on the y -axis as the best match. Data on the plot

of chosen vs. reference depth might therefore be expected to be distributed close to the line $y=x$. (Note, however, that floor and ceiling effects would tend to reduce the slope of the experimental line: subjects cannot strongly overestimate high values nor strongly underestimate low ones.) The scatter of the data is an indication of the limited precision or consistency of judgments. For loudness vibratos the slope of the line is close to 1, $M = 0.88$, $SD = 0.059$, $t(161) = -2.04$, $p = .043$, whereas for timbre vibratos it is significantly different from 1, $M = 0.82$, $SD = 0.045$, $t(128) = -3.90$, $p < .001$.

Figure 5 shows the results when participants matched vibratos of different types. The results show that when timbre vibrato depth parameter is selected to match a given loudness vibrato tone, the result has a slope close to 1, as shown as the solid black line in Figure 4, indicating that it is close to a line represented mathematically as $x=y$. This finding is based on a one-sample t -test, $M = 1.00$, being the null hypothesis value of a slope with $x = y$, $SD = 0.061$, $t(324) = 0.06$, $p < .001$, and a further t -test demonstrates that the line has an offset significantly different from 0, $M = 0.079$, $SD = 0.017$, $t(324) = 4.63$, $p < .001$. When loudness vibrato depth parameter is selected to match a given timbre vibrato tone, the results are more different: on average, the depth parameter for a loudness vibrato is 0.53 times the depth parameter of a timbre vibrato to which is judged equal to it in depth).

Combining the two observations using the geometric mean of the slopes of the two regressions: for a similar proportional change in loudness or spectral centroid, the perceived depth is 1.34 times larger in a loudness vibrato than in a timbre vibrato (average $R^2 = .44$).

Discussion

The depth matching experiment provides some quantitative information about the perception of the two different types of vibrato. First, it is clear from Figures 4 and 5 that listeners have a notion of vibrato depth: in both loudness and timbre vibrato types, higher depths are consistently matched to higher depths, with 44% of the variance in response explained. When matching vibratos of the same type, on average listeners match the target tone to 85% the depth of the reference tone.

When matching the depth of vibratos of different type, a similar result applies: listeners are able to associate a depth level of one type of vibrato to that of the other type. The depth parameter of a timbre vibrato (measured as relative change in the spectral centroid) has to be about 1.34 times higher than that of a loudness

vibrato (measured as relative change in loudness) for the two to have similar perceptual depths.

In a previous study of steady tones (Almeida et al., 2017), we found that the brightness of a tone is doubled if the spectral centroid is increased by a factor of 1.6 for tones such as those used here. So, the proportional change in spectral centroid underestimates the proportional change in brightness by a factor of about $2.0/1.6 = 1.25$. Consequently, extrapolating the salience of brightness from steady tones to vibrato, we could say that, for a judgment of similar extent of vibrato depth, the brightness variation is approximately the same as the loudness variation. Arithmetically, the brightness variation need be $1.34/(2.0/1.6) \sim 1.07$ times larger than the loudness variation, but the 7% difference is not greater than the measurement scatter.

Regression lines in Figure 5 give an average intercept of .072. This may be attributed to the interaction of floor (and ceiling) effects with statistical variations. Below a certain vibrato threshold value, a sound will be perceived as not fluctuating. Participants would choose any sound below the threshold value with equal probability, so that the average response is expected to be around half the threshold value. The threshold was estimated to be 0.15. This could explain why the slope of these linear fits is unexpectedly less than 1.

The error in matching depth for vibratos of the same type scales roughly with the depth of the reference tone (with the exception of timbre vibratos matched against their own type). This can be seen by plotting the residuals of the regression (Figure 6).

INTER-PARTICIPANT VARIABILITY

Students participating in this study had a range of musical experience. They ranked themselves into one of five categories: no musical experience (34 participants), music student (32), amateur musician (74), regular musician (17), professional musician (8). In order to test for the effect of music experience, we reduced the music experience categories to two: nonmusicians (no musical experience) and experienced musicians (regular and professional combined). We ran *t*-tests comparing the dependent variables slope and intercept of the linear fit for each of the fits, with the independent variable music experience; there was no significant difference at the $p = .01$ level.

Experiment 2. Similarity of Vibrato Types

AIMS

The aim of Experiment 2 was to compare the perception of test tones of different types, but similar depth. Would

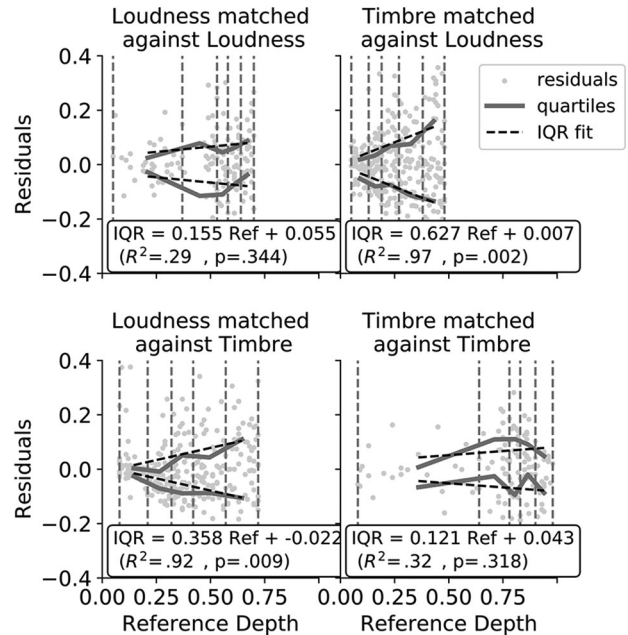


FIGURE 6. Residual plot from the data in Figures 5 and 6. The regression line is subtracted from the data which are then binned in 5 bins (delimited by the vertical dashed lines) having equal numbers of data. For each bin, the first and third quartile of the distribution of points in that bin are plotted. These graphs show how the spread varies with reference depth. For larger vibrato depths, the deviation in the matched depth appears to scale linearly with depth, with a coefficient of about 25% of the chosen depth for different type vibratos. At low depth, there appears to be a threshold in the spread. The equation in each figure represents a fit (dashed lines) of the interquartile difference (difference between the 3rd and 1st quartile) against the reference depth.

the vibrato having the same *type* as the reference tone be considered more similar, regardless of the difference in depth between the reference versus the two test tones? Or would depth be an important factor?

METHOD

Participants. Participants were the same as in Experiment 1.

Materials. Stimuli used were the same as in Experiment 1.

Procedure. Participants were asked to choose which of two tones having vibrato of different types but similar depth matched a tone whose depth was not necessarily similar. The reference tone (Tone 1) was randomly chosen to be either a loudness or timbre vibrato type with a depth parameter of 0.2, 0.3, 0.4 or 0.5 (also randomly chosen). The depth values have the same meaning as in the depth matching task in Experiment 1 for both types

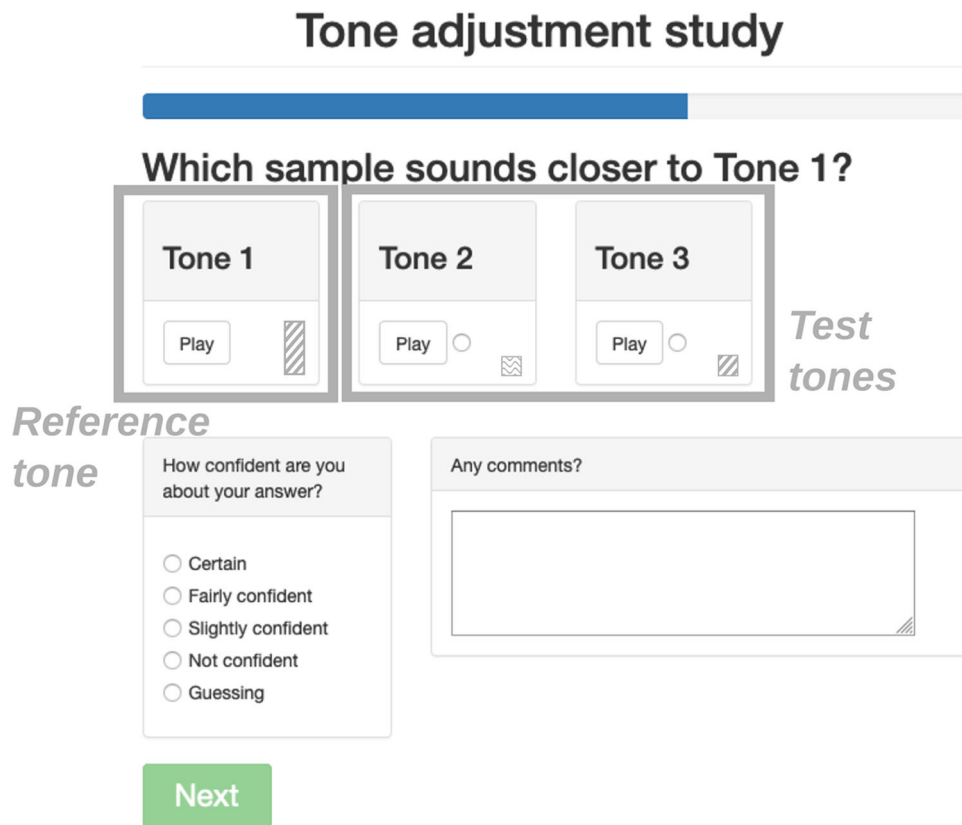


FIGURE 7. Screenshot of the interface for Experiment 2 with annotations, including bars, added in grey. Tone 1 is the reference tone and tones 2 and 3 are the test tones from which the participant has to select the best match. Small rectangles (not displayed in the experiment) show possible types (fill pattern) and depths (height) of the tones: Tones 2 and 3 always have the same depth parameter but are different types. Tone 1 can have the same depth parameter as the test tones or a different value.

of vibrato and the corresponding sounds are the same). A depth parameter of 0.5 in the loudness vibrato corresponds to a maximum loudness variation of 3.7 sones in a sound with loudness 7.5 sones. In this 500 Hz signal, a timbre vibrato depth parameter of 0.5 corresponds to a 46% change in spectral centroid: a change with amplitude 650 Hz in the average spectral centroid of 1400 Hz (see Figure 2). The test tones (Tones 2 and 3) had a randomly attributed depth parameter, the same in both tones. One was a loudness and the other was a timbre vibrato, in a random order. The participants were asked “Which sample sounds closer to tone 1?” (Figure 7).

In each trial participants were given two test tones with the same depth parameter and different type, and asked to select which was the better match with a reference tone. The vibrato type and depth of the reference tone was selected at random from 8 tones, each type having depth parameter 0.2, 0.3, 0.4, or 0.5.

RESULTS

Results are shown in Figure 8. When vibrato depths are perceived as comparable, high proportions of participants selected the test stimulus that has the same vibrato type as the reference tone. (As shown in Experiment 1, similar perceived depth means similar depth of variation in brightness and loudness, but not equal depth of variation in spectral centroid and loudness.) Figure 8 plots the fraction of best match selections where test and reference vibrato type are the same (both as a number and a shading); this fraction is plotted as a function of the vibrato depth parameters for vibrato of the reference (x) and test (y) tone. It is worth noting that this is an implicit test, because the task is to select the test tone most similar to the reference, without specifying what “similar” means. It is possible that participants sometimes group similar perceived depth as being similar, and other times similar vibrato type (loudness or timbre) as being similar.

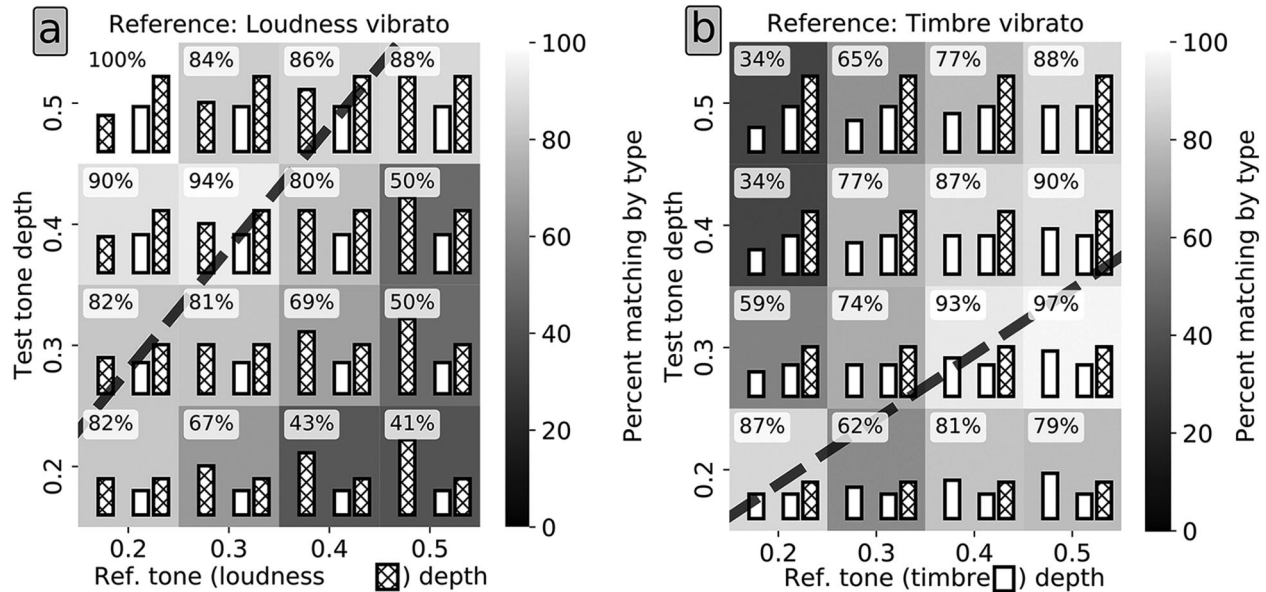


FIGURE 8. The x and y axes show the vibrato depth parameters of the reference and test tones respectively. Both the numbers and the colours represent the fraction of tones of each type matched to the same type of vibrato. The regression line of depth matching (from Experiment 1) is shown as a thick dashed line. The height of the bar at left in each cell is proportional to the perceived depth of the reference vibrato. Those of the two bars at right are proportional to the perceived depth of the two test tones. Pattern of bars shows vibrato type (hatched for loudness, plain for timbre).

Overall, participants matched test sound to a vibrato of the same type in 77% of cases. For four out of the 32 cases participants chose more often the opposite type vibrato as more similar to the test tone. These four cases were mostly those having the greatest difference in depths between test and reference. (In these four cases, 7 out of the 22 not matching by type were nonmusicians (68%), whereas only 5 (36%) out of 16 matching by type were nonmusicians). A chi-square test comparing the criterion of best match selections for the different groups showed that the difference was not significant, $\chi^2(1, N = 36) = 2.46, p = .11$.

The key to understanding these possibly surprising cases seems to be that these four are the only cases out of 32 where a shallow timbre vibrato is available to be paired with a deep loudness vibrato: either when choosing a shallow timbre vibrato to match to a deep loudness vibrato reference (bottom right of Figure 8a) or when choosing a deep loudness vibrato to match to a shallow timbre vibrato reference (top left of Figure 8b). Conversely, however, when choosing a deep timbre vibrato to match to a shallow loudness vibrato reference (top left of Figure 8a) or when choosing a shallow loudness vibrato to match to a deep timbre vibrato reference (bottom right of Figure 8b), participants match tones according to type in a quite high proportion of cases.

Discussion

As Figure 5 shows, the perceptions of the strength of vibrato depth parameter and of loudness depth parameter are similar, but not identical. For that reason, the regression lines fitted to the data of Figure 5 have been overlaid on Figure 8. The overlays show the average depth of the vibrato type that, as a single test tone, matches a dissimilar reference type. It can be observed that, along these lines, participants matched to the same type of vibrato more often than in many of the other cases (along these lines the shading is darker than in squares further from the lines). One interpretation is thus that when two dissimilar vibrato types are perceived as having similar depth it is easier to distinguish their type, and that in this case the timbre-loudness distinction becomes the more important component of dissimilarity.

Further from the lines of equal perceived depth, the rates of matching according to vibrato type drop, and at the furthest cells that were tested, participants roughly chose equally often the tone that matched one type or the other tone. This can be seen in Figure 9, which plots the distribution of best match selection as a function of the minimum perceived depth difference between test and reference.

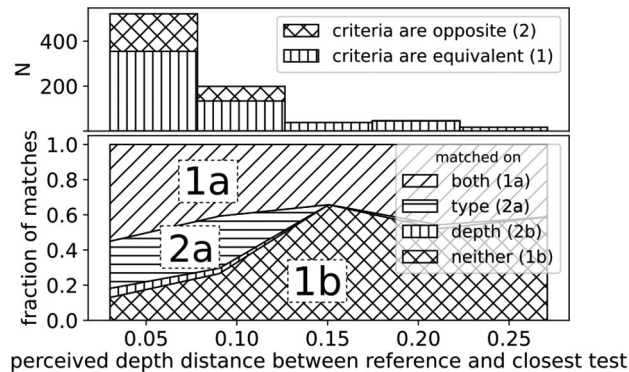


FIGURE 9. In each graph, the x -axis shows the lesser of the two possible differences in depth between the reference and the two test tones. The top graph shows the number of trials in each bin of the lower graph, as well as the number of cases where best match selection based on depth distance and based on type of vibrato would give the same response. The lower graph shows the fraction of best match selections made based on the type of vibrato or nearest perceived depth (according to the regression lines from Figure 5). The areas labeled “both” and “neither” corresponds to the top graph cases marked as “criteria are equivalent”: selecting based on type or depth would provide the same answer. It shows that, when there was a difference in best match selection according to depth or type, participants selected mostly based on the type of vibrato, except when both test tones had depths that were most different from the reference tone (above a depth difference of 0.15—right side of horizontal shading). When the depth difference was large (right half of graph), participants were equally likely to make a selection based on type and depth, or to make the “wrong” or unexpected selection (when both criteria would lead to the same best match selection).

DISTRIBUTION OF BEST MATCH SELECTION AS A FUNCTION OF PERCEIVED DEPTH DIFFERENCE

Using the results from Experiment 1, the perceived depth of each of the test and reference tones can be estimated, giving a distance in perceived depth between each of the test tones and the reference tone. Figure 9 groups best match selection in Experiment 2 as a function of the difference between the perceived depth of the test tone and that, in a given trial, of the test tone that is closer in depth to the reference. Best match selections can be classified in two groups of two cases each:

1. If the test tone that was closer in perceived depth to the reference tone also had the same type as the reference tone, the participant could have selected either that tone or the one more distant in perceived depth, so that the possible best match selection in this case would be classified as:
 - 1a) Selected on **both** criteria, because selecting the closest tone means that the subject

selected the tone with the closest perceived vibrato but also with the same type of vibrato

- 1b) Selected on **neither** criteria, because the user selected the test tone that was further in perceived depth distance and that was also the one having different type from the reference (in this group there may be tones that are almost similar in depth, so users might be choosing at random)

2. If the closer tone was a different type of vibrato from the reference (then the other test tone was of the same type as the reference), the best match selection would be classified as:

- 2a) Selected on vibrato **type**, if the user selected the tone with the same type, even though its perceived depth was further from the reference
- 2b) Selected on vibrato **depth**, if the user selected the tone that was closer in perceived depth to the reference in spite of being of different type from the source

Note that cases 1a, 1b, 2a, or 2b cover all possibilities. Also, the largest values of perceived depth differences (of the closer tone in depth) are only achieved for a small number of trials in the lower right corner of Figure 8a (because loudness vibrato of a defined depth parameter has a larger perceived depth than the same depth parameter in timbre vibrato). In this corner, choosing the closer vibrato in depth also means choosing the same type of vibrato, so for large perceived depth differences there are only trials in Group 1. Also, for these trials the test tones have very shallow vibrato whereas the reference has a deep vibrato, so it may be that subjects are often choosing at random between the two test tones.

On the left hand side of Figure 9 (small depth differences), there are large numbers of trials in both groups of cases (1 and 2) (shown in the top panel of Figure 9). Here one sees that, when both selection criteria are equivalent (1a and 1b) about 80% of subjects chose the tone that is closer to and the same type as the reference. This is as naïvely expected. Nevertheless, a substantial minority falls into the 1b case. One possible explanation is that the two test tones are not easily distinguished for at least some subjects. Case 1b is discussed further below.

Cases 2a and 2b (criteria not equivalent) show an interesting result. Here, selecting according to same type (2a) gives a different result compared to selecting according to closest depth. For the group of cases 2, a majority of subjects selected matching types of vibrato,

even though that meant selecting the vibrato with a larger depth difference from the reference. Thus, for small differences in the predicted perceived depth (left side of Figure 9), the type of vibrato is a more important criterion for associating two vibrato tones than depth. (Unfortunately, the experiment was not designed to sample this space [depth distance to reference and depth distance between test tones] evenly, so that it is hard to extend this conclusion to larger depth distances. A new experiment could be designed so as to have a larger number of tones with large depth distances falling into cases 2a and 2b.)

In some trials, particularly in the top left of Figure 8b, a majority of participants matched tones that were neither of the same type nor of similar depth (Case 1b). A possible explanation is that weak timbre vibrato seems like generic vibrato, whereas in deep timbre vibrato, the effect of spectral variation is salient. Asked to match weak timbre vibrato to either deep timbre vibrato or deep loudness vibrato, most listeners matched the weak (hypothetically sub-threshold) vibrato to the strong vibrato that does not include a spectral effect, i.e., they chose the loudness vibrato. To the listener, the (weak, non-salient) timbre vibrato and the (strong) loudness vibrato seem to be the same vibrato type: neither reveals the timbre effect. (Readers may refer to the sound tone samples used in the test, available in the supplementary materials, also at <http://newt.phys.unsw.edu.au/jw/vib1.html>). This is one of the reasons for the vanishing fraction of type-matched responses on the right hand of Figure 9.

It should also be stressed that the overlays in Figure 8 show average behavior: for individuals, the perceptual weightings differ (as the results of Experiment 1 show), and this could explain some of the spread in Figure 8. A surprising result is observed in the eight examples lying exactly on the $y=x$ diagonal. In each of these cases, one of the test tones was identical to the reference. A substantial majority (84%) of participants picked the identical tones as the most similar pair, but not all. This suggests that, for the remaining 16% of participants (of whom 10 out of 14 are nonmusicians), two vibratos of similar depth but different type are chosen as a match.

General Discussion and Conclusions

Participants are reasonably sensitive to vibrato depth and able to match the depths of two tones with the same vibrato type: a linear relation explains 62% (when matching loudness vibrato) or 72% (when matching timbre) of the variance. A linear relation only explains 41% of variance when matching the vibrato depth across types. The linear relation's positive intercept suggests that participants cannot detect fluctuations of less than about 15% either in loudness or spectral centroid. When matching across vibrato types, the linear relationship suggests that proportional fluctuations in loudness are perceived as being about 1.3 times deeper than proportional fluctuations in spectral centroid and thus (by extrapolating from previous measurements on steady tones), about the same depth as proportional changes in brightness.

Recognition of vibrato similarity is complicated: when participants perceived their depths to be similar, about 80% associate tones having similar vibrato type. However, when two vibrato tones have very different depths, only about 50% of participants associate the tones by type, a sign that they might be answering randomly in these cases. When timbre vibrato depth was sufficiently small, participants associated it more often with a deep loudness vibrato than a deep timbre vibrato, suggesting that, for many people, spectral effects are not perceptually salient when the variation in spectral centroid is about 30% of the mean value (350 Hz, in the current examples) or less.

So despite fairly linear sensitivity to vibrato depth regardless of vibrato type, some poorly understood findings between physical signal and perception of timbre vibrato remain, and more research is needed in timbre vibrato perception, which, in distinction to pitch and loudness vibratos has been considerably neglected.

Author Note

We thank the Australian Research Council for supporting this project, and Professor John R. Smith for helpful discussions.

Correspondence concerning this article should be addressed to Andre Almeida, School of Physics, UNSW, Sydney 2052, Australia. E-mail: a.almeida@unsw.edu.au

References

- ALMEIDA, A., SCHUBERT, E., SMITH, J., & WOLFE, J. (2017). Brightness scaling of periodic tones. *Attention, Perception, and Psychophysics*, 79, 1892–1896.
- BENADE, A. H. (1976). *Fundamentals of musical acoustics*. New York: Oxford University Press.

- CABRERA, D. (1999). PSYSOUND: A computer program for psychoacoustical analysis. In *Proceedings of the Australian Acoustical Society Conference* (Vol. 24, pp. 47–54). Melbourne, Australia: AASC.
- COLEMAN, R. F. (1979). Acoustic and perceptual factors in vibrato. In V. Lawrence & B. Weinberg (Eds.), *Transcripts of the Eighth Symposium on Care of The Professional Voice* (Part I, 36–38). New York.
- DESAIN, P., HONING, H., AARTS, R., TIMMERS, R & WINDSOR W. L. (1999). Rhythmic aspects of vibrato. In P. Desain & W. L. Windsor (Eds.), *Rhythm perception and production* (pp. 203–216). Lisse: Swets & Zeitlinger.
- GERINGER, J. M., MACLEOD, R. B., MADSEN, C. K., & NAPOLES, J. (2014). Perception of melodic intonation in performances with and without vibrato. *Psychology of Music*, 43, 675–685. <https://doi.org/10.1177/0305735614534004>
- GOUGH, C. (2005) Measurement, modelling and synthesis of violin vibrato sounds. *Acta Acustica United With Acustica*, 91, 229–240
- HAJDA, J. M. (1999). *The effect of time-variant acoustical properties on orchestral instrument timbres* (PhD thesis). University of California, Los Angeles.
- KATOK, D. (2016). *The versatile singer: A guide to vibrato and straight tone* (PhD thesis). City University of New York. Academic Works. https://academicworks.cuny.edu/gc_etds/1394
- LEEK, M. R. (2001). Adaptive procedures in psychophysical research. *Perception and Psychophysics*, 63, 1279–1292.
- PAPICH, G., & RAINBOW, E. (1974). A pilot study of performance practices of twentieth-century musicians. *Journal of Research in Music Education*, 22, 24–34.
- PRAME, E. (1997). Vibrato extent and intonation in professional Western lyric singing. *Journal of the Acoustical Society of America*, 102, 616–621.
- SEASHORE, C. E. (1937) The psychology of music. VI. The vibrato: (1) What is it? *Music Educators Journal*, 23, 30–32.
- SEASHORE, C. E. (1938). *Psychology of music*. New York: McGraw-Hill.
- SUNDBERG, J. (1977). Research in singing. *Journal of Studies of the Soprano Voice*, 1, 25–35.
- ZHANG, M., BOCKO, M., BEAUCHAMP, J. (2015). Measurement and analysis of musical vibrato parameters. *Proceedings of Meetings in Acoustics*, 23, 035004. DOI: 10.1121/2.0000136