A sliding two-alternative forced-choice paradigm for pitch discrimination

Dorothée Arzounian,[1] Mathilde de Kerangal,[2] and Alain de Cheveigné[1,a]

[1]*Laboratoire des Systèmes Perceptifs, Département d'Études Cognitives, École Normale Supérieure, PSL Research University, CNRS, 29 rue d'Ulm, F-75230 Paris, France*

[2]*The Ear Institute, University College London, 332 Grays Inn Road, Kings Cross, London WC1X 8EE, United Kingdom*

Studies that measure frequency discrimination often use 2, 3, or 4 tones per trial. This paper shows an investigation of a two-alternative forced choice (2AFC) task in which each tone of a series is judged relative to the previous tone ("sliding 2AFC"). Potential advantages are a greater yield (number of responses per unit time), and a more uniform history of stimulation for the study of context effects, or to relate time-varying performance to cortical activity. The new task was evaluated relative to a classic 2-tone-per-trial 2AFC task with similar stimulus parameters. For each task, conditions with different stimulus parameters were compared. The main results were as follows: (1) thresholds did not differ significantly between tasks when similar parameters were used. (2) Thresholds did differ between conditions for the new task, showing a deleterious effect of inserting relatively large steps in the frequency sequence. (3) Thresholds also differed between conditions for the classic task, showing an advantage for a fixed frequency standard. There was no indication that results were more variable with either task, and no reason was found not to use the new sliding 2AFC task in lieu of the classic 2-tone-per-trial 2AFC task. © 2017 Acoustical Society of America.
[http://dx.doi.org/10.1121/1.4992030]

[JJL]   Pages: 167–172

I. INTRODUCTION

Pitch is a prominent quality of sound important for music and speech, the object of keen interest since antiquity (de Cheveigné, 2005; Plack, 2010; McDermott and Oxenham, 2008; Oxenham, 2012). Pitch discrimination has been explored in many studies, to characterize perceptual and sensory limits (Moore, 1973; Bernstein and Oxenham, 2003; Demany *et al.*, 2009; Oxenham *et al.*, 2011; Micheyl *et al.*, 2012; Micheyl *et al.*, 2010a) in the psychophysical tradition, to explore effects of context or memory (Matthews and Stewart, 2008; Ries and DiGiovanni, 2009; Nahum *et al.*, 2010; Raviv *et al.*, 2012; Mathias *et al.*, 2010; Micheyl *et al.*, 2010b), to characterize interindividual differences (Semal and Demany, 2006; Mathias *et al.*, 2010) and to relate them to factors such as musicianship (Kishon-Rabin *et al.*, 2001; Nikjeh *et al.*, 2008; Peretz *et al.*, 2002; Micheyl *et al.*, 2006; Tervaniemi *et al.*, 2005; Bianchi *et al.*, 2016), intellectual abilities or impairment (Acton and Schroeder, 2001; Parsons *et al.*, 2009), effects of training and plasticity (Demany and Semal, 2002; Carcagno and Plack, 2011b; Micheyl *et al.*, 2006; Goldsworthy and Shannon, 2014; Carcagno and Plack, 2011a), or electrophysiological responses (Nikjeh *et al.*, 2008; Barker *et al.*, 2011; Carcagno and Plack, 2011a; Petacchi *et al.*, 2011; Clinard *et al.*, 2010; Tervaniemi *et al.*, 2005). A common observation is large interindividual differences, with thresholds ranging from about 0.2% for trained subjects and musicians to greater than 10% for naive subjects. Marked training effects are also reported.

Discrimination thresholds can be measured using various methods. Adaptive procedures are usually preferred over constant stimuli as they ensure a dense sampling of the psychometric curve in the region of the threshold. That region is hard to situate prior to the experiment because of interindividual differences and training effects. In one common procedure (two-interval two-answer forced choice, 2I-2AFC) two tones are presented on each trial with slightly different frequencies, and the subject is required to answer which one has a higher pitch. In a variant of this procedure, the tones have either the same or a different frequency, and the subject is required to judge whether the pitch of the two tones is the same or different. The frequency difference is varied adaptively to converge on the threshold (usually corresponding to 70.7% or 79% correct). Other studies have used 3 tones (Bernstein and Oxenham, 2003; Kishon-Rabin *et al.*, 2001; Nikjeh *et al.*, 2008; Bianchi *et al.*, 2016) or 4 tones (Semal and Demany, 2006; Mathias *et al.*, 2010; Amitay *et al.*, 2006) per trial.

In this paper, we investigate a one-tone-per-trial "sliding 2AFC" paradigm in which the subject is presented with a sequence of tones and asked, after each tone, to judge whether its pitch is higher or lower than that of the preceding tone. A first motivation is to improve the yield (number of useful judgments per unit time) relative to classic paradigms, so as to reduce experimental time or increase statistical power for threshold estimates. This is particularly useful in

a)Also at: The Ear Institute, University College London, 332 Grays Inn Road, Kings Cross, London WC1X 8EE, United Kingdom. Electronic mail: alain.de.cheveigne@ens.fr

experiments that require a large number of trials (e.g., Demany and Semal, 2002). A second motivation is to facilitate the study of the effect of prior context on pitch judgments. It is well known that trial-to-trial roving of the frequency of the first tone of each pair in tasks involving 2 tones per trial increases thresholds (Bull and Cuddy, 1972; Ahissar et al., 2006; Matthews and Stewart, 2008), reflecting a more general sensitivity of pitch judgments to the history of prior tone frequencies (Ruusuvirta, 2000; Raviv et al., 2012). Studying such effects should be easier with a continuous paradigm, in which all tones play the same role, and all are followed by a judgment, than in 2-, 3-, or 4-interval tasks in which successive tones play different roles.

The primary aim of the present study is to check whether the new procedure introduces any unexpected artifacts, such as increased bias or variability. For this we run both procedures with comparable parameters on the same subjects, in interleaved blocks to control for effects of time and practice on performance. A desirable outcome, relative to our goals, would be a lack of difference between the thresholds measured with the new and classic procedures, however a lack of effect is hard to demonstrate statistically (Kriegeskorte et al., 2009). For that reason, we also included conditions for which we *do* expect different thresholds, to establish whether the experiment is sufficiently sensitive to reveal the hypothesized artifacts, should they exist. Thus, we ran five different conditions, three with a classic two-interval 2AFC procedure, and two with the new sliding 2AFC procedure.

II. METHODS

The principal aim of these experiments is to validate a continuous procedure to measure pitch discrimination performance, and compare it to a standard procedure using 2 tones per trial. In order to achieve this goal, five conditions were presented, distributed over 3 sessions.

A. Subjects

Fifteen subjects took part in the experiments, 10 women and 5 men aged from 21 to 27 years. Seven of them were musicians. Subjects who had played an instrument for at least 5 years and still had a regular practice were qualified as musicians. All musician subjects were members or former members of a music school or conservatory. One non-musician woman was excluded from the analysis because of her abnormally poor performance on the tasks.

Participants had no history of audiological or neurological disorders, and reported normal hearing which was confirmed by an audiogram in the frequency band used in the experiments. The procedures were approved by the CERES ethics committee (IRB 20131100001072), the participants were informed of the procedures and a written consent was obtained from each of them. All subjects performed all three sessions.

B. Stimuli

Stimuli consisted of 100 ms pure tones with 10 ms raised cosine onset and offset ramps. For the sliding one-tone-per-trial

procedure (conditions O1 and O2) a response was requested after each tone. For the two-tone procedure (conditions T1, T2, T3), a response was requested after each pair of tones presented with a 500 ms stimulus onset asynchrony. In both procedures, the response triggered the onset of a new tone starting the next trial after a 500 ms interval. In the one-tone procedure, the frequency varied from one tone to the next with random direction, starting from $f_0 = 1000$ Hz. An unbiased random walk can produce extreme frequency values; to avoid this situation the probability of an up transition (0.5 at 1000 Hz) decreased linearly in frequency by a factor 0.0003/Hz, so that frequency remained near 1000 Hz. In this region, up and down transitions were approximately equiprobable (see Sec. III). In both procedures, the magnitude of the frequency step between the tones to be compared was determined according to a weighted one-down algorithm to track the 75% correct point on the psychometric function (Kaernbach, 1991). After every wrong answer the relative frequency difference $|\Delta f/f|$ was multiplied by 2, after every correct answer it was divided by $2^{1/3}$. For each experimental block the step size was initially set to 10% and adapted with a single adaptive track (for rapid convergence) until the first wrong answer after at least 12 trials, after which the procedure switched to 4 interleaved independent tracks (Leek et al., 1991). The size of step j was adjusted based on the response to trial $j - 4$. The interleaved tracks were used to reduce serial correlation between successive frequency step sizes, for the purpose of analysis in a future study.

Stimuli were synthesized at a sampling rate of 44.1 kHz and outputted via a Meridian Explorer2 sound card with 16 bit resolution. Sounds were delivered diotically through Beyerdynamic DT 770 pro 250 ohm headphones, at a listening level of 70 dB sound pressure level. Listeners were tested individually in a double-walled sound attenuating booth. Instructions were provided verbally, and displayed again on a graphic interface. Subjects used the keyboard to answer, and received visual feedback (green for correct vs red for incorrect) after each response.

C. Procedures and conditions

For the sliding one-tone procedure (conditions O1, O2), the subject had to report after each tone whether it was higher or lower than the previous tone (upper right panel of Fig. 1). In condition O1 the tone-to-tone frequency step size was determined by the adaptive procedure. In condition O2 one step out of two was determined by the adaptive procedure and the other was set to 6% (one semitone). The response after that step was not taken into account for the adaptive procedure.

For the two-tone-per-trial procedure (conditions T1, T2, T3) the subject had to report after each tone pair whether the second tone was higher or lower than the first. The within-pair step size was determined by the adaptive procedure, but conditions T1, T2, and T3 differed in how the between-pair step size was determined. In condition T1 it was determined by the random procedure, leading to a similar stimulus sequence pattern as in O1. In condition T2 the between-pair step size was zero (the first tone of a pair had the same
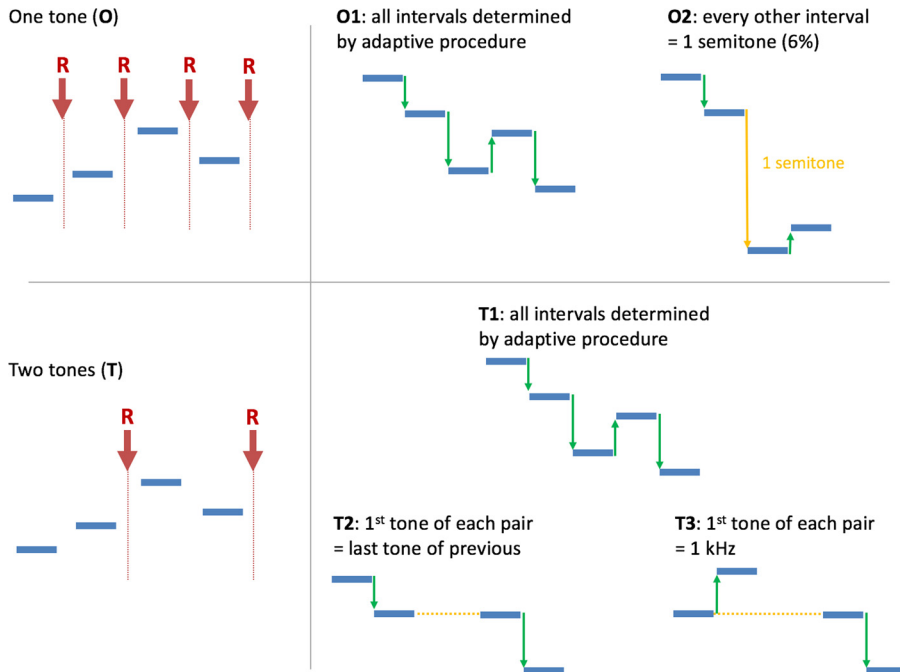
FIG. 1. (Color online) (Top left) Illustration of the sliding one-tone-per-trial procedure. Subjects had to give an answer (pitch went up or down) after each tone. (Bottom left) Two-tone-per-trial procedure. Subjects had to give an answer after each pair of tones. (Top right) Two conditions that used the one-tone procedure. For O1 all intervals were determined by the adaptive procedure, for O2 one interval out of two was set to ± 1 semitone. (Bottom right) Three conditions that used the two-tone procedure. For T1 all intervals (within and between pairs) were determined by the adaptive procedure (as in O1). For T2 the first tone of each pair had the same frequency as the second tone of the preceding pair. For T3 the first tone of each pair was 1 kHz.

frequency as the second tone of the preceding pair). In condition T3 the frequency of the first tone was always 1 kHz (fixed standard; lower right panel of Fig. 1).

Conditions O1 and T1 share very similar tone sequences, and are expected to yield the same thresholds in the absence of an artifact specific to the one- or two-tone procedure. Condition O2 is expected to yield larger thresholds than O1 because small near-threshold steps are interspersed with relatively large steps (6%, a semitone). Condition T2 is expected to yield smaller thresholds than T1 because the first tone of a pair is, in effect, repeated twice. Condition T3 is also expected to yield smaller thresholds than other conditions because the first tone of each pair (standard) is the same.

The experiment was carried out in three 1-hour sessions (including instructions and training) happening on different days. Two conditions were presented in each session, interleaved into six blocks to control for effects of time and practice (for example, due to learning or fatigue).

In each block, the subject performed one condition. In the first session, the subjects performed the main conditions O1 and T1, in the second session they performed conditions T2 and T3, and in the third and last session, they performed conditions O1 and O2. 120 trials were presented per block, except for condition O2 where 240 trials were presented per block (responses to 6% steps were ignored in the analysis). Each session started with two short training blocks (60 trials each, 120 for O2), one for each condition in that session. This was to familiarize the subjects with the procedures and ensure that they understood what feature of the sound they had to listen to.

D. Analysis

A psychometric function of the form

$$P = 0.5 + 0.5 \frac{1}{1 + e^{-s(x-x_0)}},$$

where P is the probability of a correct answer and $x = \log(|\Delta f/f|)$, was fit to the data by adjusting the two parameters s and x_0, using a least-squares error method. Δf refers to the difference between the frequencies of the two tones to be compared, and f is (arbitrarily) the lowest of these two frequencies. Figure 2 (top) shows response data of an experimental run together with the fit. The histograms in Fig. 2 (bottom) indicate the distribution of values sampled by the adaptive procedure.

III. RESULTS

Fourteen paid subjects performed three sessions of six blocks each. In each session, two conditions were interleaved for a total of 360 responses per condition (720 for
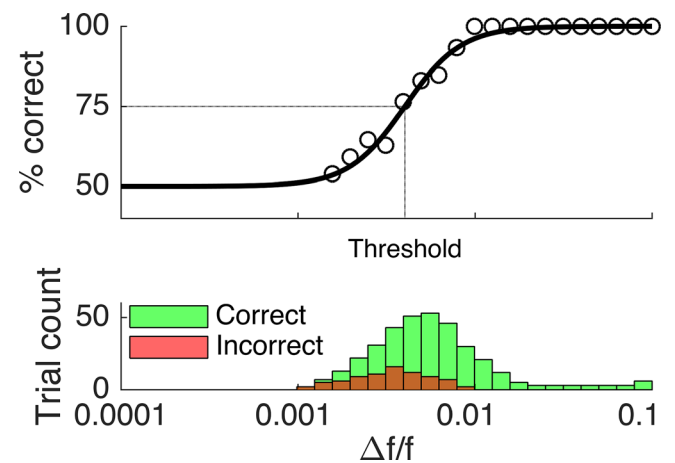

FIG. 2. (Color online) (Top) Psychometric function fit (line) to experimental data (circles) for a typical experimental run. The threshold is defined as the abscissa at 75% correct on the psychometric function. (Bottom) Histogram of values of relative frequency differences $\Delta f/f$ sampled by the adaptive procedure. Colors distinguish trials for which the response was correct (green) or incorrect (red).

J. Acoust. Soc. Am. **142** (1), July 2017

Arzounian *et al.* 169

condition O2 of which 360 were retained for analysis). Condition O1 was presented twice over two sessions.

Individual results are shown in Fig. 3. The thresholds in condition O1 are presented separately for session 1 (O1a) and session 3 (O1b). The range of threshold values is consistent with the literature (Micheyl *et al.*, 2006), although one study reported considerably larger values (Raviv *et al.*, 2012). The lowest thresholds found here (~0.2%) are comparable to those reported for trained subjects in classic papers on pitch discrimination (Moore, 1973; Micheyl *et al.*, 2006; Demany and Semal, 2002).

A repeated measure analysis of variance with factors repetition, conditions and groups demonstrated significant main effects of condition [Greenhouse-Geisser correction $F(2.43, 29.21) = 33.7$, $p < 0.001$] and group [$F(1, 12) = 8.35$, $p = 0.014$], but no main effect of repetition [$F(2, 24) = 2.08$, $p = 0.15$]. Interactions between repetition and condition [$F(10, 120) = 0.85$, $p = 0.58$] and between condition and musicianship [$F(5, 60) = 1.78$, $p = 0.13$] were not significant. The interaction between repetition and musicianship was marginally significant [$F(2, 24) = 3.63$, $p = 0.042$], reflecting a slight tendency for performance to improve over sessions for non-musicians (Fig. 3). The three-way interaction was not significant. As effects involving repetition were either non-significant or small, data were merged over blocks to produce one threshold per condition and per subject. Group-mean thresholds per condition are shown in Fig. 4, error bars indicating standard error of the mean.

Focusing on pairwise comparisons, of main interest to our study we found no threshold difference between O1a and T1 (session 1), even in the absence of Bonferroni correction for multiple comparisons ($p > 0.05$, sign test). The lack of a significant effect carries little weight in itself, as it could result for example from a lack of power (Kriegeskorte *et al.*, 2009). In this respect, it is of interest that other differences between conditions were found to be significant and of considerable magnitude.

After Bonferroni correction for 6 comparisons, the difference between conditions O1b and O2 in session 3 was significant ($p = 0.011$, corrected), thresholds for O2 being larger than for O1b, suggesting a deleterious effect of inserting a

FIG. 4. Frequency discrimination thresholds geometrically averaged over subjects and blocks for each condition. Error-bars represent standard error of the mean. Stars indicate significant differences as assessed by sign tests with Bonferroni correction (c = 6) for multiple comparisons (* for $p < 0.05$, *** for $p < 0.001$), "n.s." indicates the difference is not significant even before Bonferroni correction.

relatively large frequency step (6%) prior to an interval to be judged. There was a significant threshold difference between T1 and T3 ($p < 0.001$, corrected), thresholds for T3 being lower than for T1, indicating better performance when the first tone of all pairs was set to the same frequency (fixed standard) rather than roving. Thresholds were also lower for T3 than T2 ($p = 0.011$, corrected). The difference between T1 and T2 was significant before Bonferroni correction ($p = 0.013$) but not after correction. We found no significant difference between O1a and O1b ($p > 0.05$, uncorrected), again failing to demonstrate a change of performance over time. There were no differences in psychometric function slopes between any conditions.

To summarize, we found no major difference between procedures (sliding one-tone-per-trial and classic two-tone-per-trial), whereas for both procedures we found significant effects of manipulating the prior context (O1 vs O2, T1 vs T3 and T2 vs T3). There was no evidence of greater variability (over trials or subjects) or learning effects for one procedure relative to the other.

IV. DISCUSSION

The primary aim of this study was to compare a one-tone-per-trial sliding 2AFC discrimination threshold measurement procedure to a standard two-tone-per-trial 2AFC procedure. We found that both procedures were sensitive to relatively subtle manipulations of stimulus context, but yielded equivalent thresholds when applied to comparable stimulus conditions. Subjects reported no difficulty with either task and, overall, we found no indication of a problem specific to the new procedure relative to the old. Nonetheless, the lowest thresholds were found with the two-tone-per trial procedure with fixed reference (T3).

A secondary aim of this study was to compare conditions that differ in the pattern of prior context, using the same subjects and controlling for within-session and between-session serial effects so as to maximize sensitivity to potentially small effects. A common observation in the
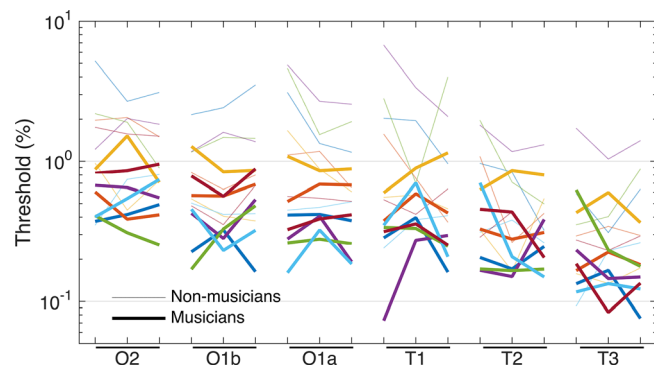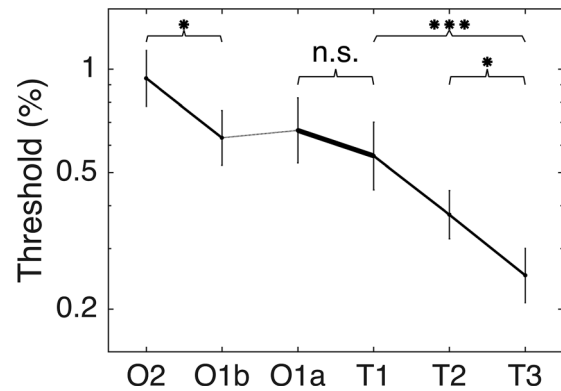

FIG. 3. (Color online) Individual frequency discrimination thresholds for each condition and each block. Thresholds of musicians are plotted as thick lines, those of non-musicians as thin lines. Each participant is associated with a color (online version). Within each condition (horizontal axis), thresholds obtained in the 3 blocks are plotted in chronological order.

literature is that thresholds are larger if the frequency of the first tone of each pair is *roved* rather than constant (fixed standard) (Bull and Cuddy, 1972; Ahissar *et al.*, 2006; Matthews and Stewart, 2008). In our study, every condition involved some form of "roving" with the exception of T3. That condition yielded the lowest thresholds, consistent with prior studies. Those studies considered roving over wide frequency ranges, our study extends their conclusions to a roving range on the order of a few times the discrimination threshold. Thresholds in our study increased with the amplitude of roving (quantified as the average interval from one "standard" to the next), the highest being found for ∼6% (O2) and the lowest for 0% (T3).

The deleterious effect of roving has been interpreted as reflecting an inability to form a reliable perceptual anchor (Berliner and Durlach, 1973), or greater requirements placed on working memory (Zhang *et al.*, 2016), or more general informational masking or "confusion" effects (Mathias *et al.*, 2010). Those accounts invoke properties of the stimulus set as a whole; an alternative is that the internal representation of the standard on any trial is influenced by the immediately preceding tone or tones. In a companion study, we found that context effects were well predicted by a linear model including (in its simplest form) a term proportional to the frequency of the immediately preceding tone. As the influence of an irrelevant stimulus feature on decisions must lead to suboptimal performance, a stronger influence in the case of roving might indeed lead to a higher threshold.

Subjects did not report any particular difficulty with either procedure. Subjectively, the sliding 2AFC procedure is straightforward and refreshingly brisk, as each stimulus event calls for a response. The task is analogous to perception of music, where each tone of a melody anchors both the intervals that precede and follow it. Thresholds did not differ from a standard 2-tone-per-trial 2AFC procedure with similar stimuli. This is a useful outcome, as the sliding procedure is considerably more efficient in terms of number of judgments per unit time. The greater density of judgments allows fluctuations in performance (for example, related to learning, fatigue or brain state changes) to be followed with finer temporal resolution. All tones play the same role (each tone serves both as a test tone relative to the previous tone, and as a reference relative to the following tone) which is useful when studying the effects of the extended tonal history preceding each judgment. As described the procedure required subjects to give up-vs-down judgments, but it could presumably be adapted as a same-different task, for example to cater to listeners who have difficulties identifying the *direction* of pitch changes (Semal and Demany, 2006; Mathias *et al.*, 2010). The sliding 2AFC procedure may also be of use to measure discrimination thresholds for other stimulus features such as intensity.

ACKNOWLEDGMENTS

Acton, G. S., and Schroeder, D. H. (**2001**). "Sensory discrimination as related to general intelligence," Intelligence **29**(3), 263–271.

Ahissar, M., Lubin, Y., Putter-Katz, H., and Banai, K. (**2006**). "Dyslexia and the failure to form a perceptual anchor," Nat. Neurosci. **9**(12), 1558–1564.

Amitay, S., Irwin, A., Hawkey, D. J. K., Cowan, J. A., and Moore, D. R. (**2006**). "A comparison of adaptive procedures for rapid and reliable threshold assessment and training in naive listeners," J. Acoust. Soc. Am. **119**(3), 1616–1625.

Barker, D., Plack, C. J., and Hall, D. A. (**2011**). "Human auditory cortical responses to pitch and to pitch strength," NeuroReport **22**(3), 111–115.

Berliner, J. E., and Durlach, N. I. (**1973**). "A perceptual anchor model for context coding in intensity perception," J. Acoust. Soc. Am. **54**(1), 336.

Bernstein, J. G., and Oxenham, A. J. (**2003**). "Pitch discrimination of diotic and dichotic tone complexes: Harmonic resolvability or harmonic number?," Acoust. Soc. Am. J. **113**(6), 3323–3334.

Bianchi, F., Santurette, S., Wendt, D., and Dau, T. (**2016**). "Pitch discrimination in musicians and non-musicians: Effects of harmonic resolvability and processing effort," J. Assoc. Res. Otolaryngol. **17**(1), 69–79.

Bull, A. R., and Cuddy, L. L. (**1972**). "Recognition memory for pitch of fixed and roving stimulus tones," Percept. Psychophys. **11**(1), 105–109.

Carcagno, S., and Plack, C. J. (**2011a**). "Subcortical plasticity following perceptual learning in a pitch discrimination task," J. Assoc. Res. Otolaryngol. **12**(1), 89–100.

Carcagno, S., and Plack, C. J. (**2011b**). "Pitch discrimination learning: Specificity for pitch and harmonic resolvability, and electrophysiological correlates," J. Assoc. Res. Otolaryngol. **12**(4), 503–517.

Clinard, C. G., Tremblay, K. L., and Krishnan, A. R. (**2010**). "Aging alters the perception and physiological representation of frequency: Evidence from human frequency-following response recordings," Hear. Res. **264**(1), 48–55.

de Cheveigné, A. (**2005**). "Pitch perception models," in *Pitch* (Springer, New York), pp. 169–233.

Demany, L., Carlyon, R. P., and Semal, C. (**2009**). "Continuous versus discrete frequency changes: Different detection mechanisms?," J. Acoust. Soc. Am. **125**(2), 1082–1090.

Demany, L., and Semal, C. (**2002**). "Learning to perceive pitch differences," Acoust. Soc. Am. J. **111**(3), 1377–1388.

Goldsworthy, R. L., and Shannon, R. V. (**2014**). "Training improves cochlear implant rate discrimination on a psychophysical task," J. Acoust. Soc. Am. **135**(1), 334–341.

Kaernbach, C. (**1991**). "Simple adaptive testing with the weighted up-down method," Atten. Percept. Psychophys. **49**(3), 227–229.

Kishon-Rabin, L., Amir, O., and Vexler, Y. (**2001**). "Pitch discrimination: Are professional musicians better than non-musicians?," J. Basic Clin. Physiol. Pharmacol. **12**, 125–143.

Kriegeskorte, N., Simmons, W. K., Bellgowan, P. S. F., and Baker, C. I. (**2009**). "Circular analysis in systems neuroscience: The dangers of double dipping," Nat. Neurosci. **12**(5), 535–540.

Leek, M. R., Hanna, T. E., and Marshall, L. (**1991**). "An interleaved tracking procedure to monitor unstable psychometric functions," J. Acoust. Soc. Am. **90**(3), 1385–1397.

Mathias, S. R., Micheyl, C., and Bailey, P. J. (**2010**). "Stimulus uncertainty and insensitivity to pitch-change direction," J. Acoust. Soc. Am. **127**(5), 3026–3037.

Matthews, W. J., and Stewart, N. (**2008**). "The effect of stimulus range on two-interval frequency discrimination," J. Acoust. Soc. Am. **123**(4), EL45–EL51.

McDermott, J. H., and Oxenham, A. J. (**2008**). "Music perception, pitch, and the auditory system," Curr. Opin. Neurobiol. **18**(4), 452–463.

Micheyl, C., Delhommeau, K., Perrot, X., and Oxenham, A. J. (**2006**). "Influence of musical and psychoacoustical training on pitch discrimination," Hear. Res. **219**(1), 36–47.

Micheyl, C., Divis, K., Wrobleski, D. M., and Oxenham, A. J. (**2010a**). "Does fundamental-frequency discrimination measure virtual pitch discrimination?," J. Acoust. Soc. Am. **128**(4), 1930–1942.

Micheyl, C., Keebler, M. V., and Oxenham, A. J. (**2010b**). "Pitch perception for mixtures of spectrally overlapping harmonic complex tones," J. Acoust. Soc. Am. **128**(1), 257–269.

Micheyl, C., Xiao, L., and Oxenham, A. J. (**2012**). "Characterizing the dependence of pure-tone frequency difference limens on frequency, duration, and level," Hear. Res. **292**(1), 1–13.

Moore, B. (**1973**). "Frequency difference limens for short-duration tones," J. Acoust. Soc. Am. **54**(3), 610–619.

Nahum, M., Daikhin, L., Lubin, Y., Cohen, Y., and Ahissar, M. (**2010**). "From comparison to classification: A cortical tool for boosting perception," J. Neurosci. **30**(3), 1128–1136.

Nikjeh, D. A., Lister, J. J., and Frisch, S. A. (**2008**). "Hearing of note: An electrophysiologic and psychoacoustic comparison of pitch discrimination between vocal and instrumental musicians," Psychophysiology **45**(6), 994–1007.

Oxenham, A. J. (**2012**). "Pitch perception," J. Neurosci. **32**(39), 13,335–13,338.

Oxenham, A. J., Micheyl, C., Keebler, M. V., Loper, A., and Santurette, S. (**2011**). "Pitch perception beyond the traditional existence region of pitch," Proc. Natl. Acad. Sci. U.S.A. **108**(18), 7629–7634.

Parsons, L. M., Petacchi, A., Schmahmann, J. D., and Bower, J. M. (**2009**). "Pitch discrimination in cerebellar patients: Evidence for a sensory deficit," Brain Res. **1303**, 84–96.

Peretz, I., Ayotte, J., Zatorre, R. J., Mehler, J., and Ahad, P. (**2002**). "Congenital amusia: A disorder of fine-grained pitch discrimination," Neuron **33**(2), 185–191.

Petacchi, A., Kaernbach, C., Ratnam, R., and Bower, J. M. (**2011**). "Increased activation of the human cerebellum during pitch discrimination: A positron emission tomography (PET) study," Hear. Res. **282**(1–2), 35–48.

Plack, C., ed. (**2010**). *Pitch Perception* (Oxford University Press, Oxford, UK).

Raviv, O., Ahissar, M., and Loewenstein, Y. (**2012**). "How recent history affects perception: The normative approach and its heuristic approximation," PLoS Comput. Biol. **8**(10), e1002731.

Ries, D. T., and DiGiovanni, J. J. (**2009**). "Effects of recurrent tonal information on auditory working memory for pitch," Hear. Res. **255**(1–2), 14–21.

Ruusuvirta, T. (**2000**). "Proactive interference of a sequence of tones in a two-tone pitch comparison task," Psychonom. Bull. Rev. **7**(2), 327–331.

Semal, C., and Demany, L. (**2006**). "Individual differences in the sensitivity to pitch direction," J. Acoust. Soc. Am. **120**, 3907–3915.

Tervaniemi, M., Just, V., Koelsch, S., Widmann, A., and Schröger, E. (**2005**). "Pitch discrimination accuracy in musicians vs nonmusicians: An event-related potential and behavioral study," Exp. Brain Res. **161**(1), 1–10.

Zhang, Y., Moore, D. R., Guiraud, J., and Molloy, K. (**2016**). "Auditory discrimination learning: Role of working memory," PloS One **11**(1), e0147320.