



Time-dependent discrimination advantages for harmonic sounds suggest efficient coding for memory

Malinda J. McPherson^{a,b,c,1} and Josh H. McDermott^{a,b,c,d,1}

^aDepartment of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139; ^bProgram in Speech and Hearing Bioscience and Technology, Harvard University, Boston, MA 02115; ^cMcGovern Institute for Brain Research, Massachusetts Institute of Technology, Cambridge, MA 02139; and ^dCenter for Brains, Minds and Machines, Massachusetts Institute of Technology, Cambridge, MA 02139

Edited by Wilson S. Geisler, The University of Texas at Austin, Austin, TX, and approved November 5, 2020 (received for review May 5, 2020)

Perceptual systems have finite memory resources and must store incoming signals in compressed formats. To explore whether representations of a sound's pitch might derive from this need for compression, we compared discrimination of harmonic and inharmonic sounds across delays. In contrast to inharmonic spectra, harmonic spectra can be summarized, and thus compressed, using their fundamental frequency (f0). Participants heard two sounds and judged which was higher. Despite being comparable for sounds presented back-to-back, discrimination was better for harmonic than inharmonic stimuli when sounds were separated in time, implicating memory representations unique to harmonic sounds. Patterns of individual differences (correlations between thresholds in different conditions) indicated that listeners use different representations depending on the time delay between sounds, directly comparing the spectra of temporally adjacent sounds, but transitioning to comparing f0s across delays. The need to store sound in memory appears to determine reliance on f0-based pitch and may explain its importance in music, in which listeners must extract relationships between notes separated in time.

psychoacoustics | short-term memory | individual differences | pitch perception

Our sensory systems transduce information at high bandwidths but have limited resources to hold this information in memory. In vision, short-term memory is believed to store schematic structure extracted from image intensities, e.g., object shape, or gist, that might be represented with fewer bits than the detailed patterns of intensity represented on the retina (1–4). For instance, at brief delays visual discrimination shows signs of being based on image intensities, believed to be represented in high-capacity but short-lasting sensory representations (5). By contrast, at longer delays more abstract (6, 7) or categorical (8) representations are implicated as the basis of short-term memory.

In other sensory modalities, the situation is less clear. Audition, for instance, is argued to also make use of both a sensory trace and a short-term memory store (9, 10), but the representational characteristics of the memory store are not well characterized. There is evidence that memory for speech includes abstracted representations of phonetic features (11, 12) or categorical representations of phonemes themselves (13–15). Beyond speech, the differences between transient and persistent representations of sound remain unclear. This situation plausibly reflects a historical tendency within hearing research to favor simple stimuli, such as sinusoidal tones, for which there is not much to abstract or compress. Such stimuli have been used to characterize the decay characteristics of auditory memory (16–19), its vulnerability to interference (20, 21), and the possibility of distinct memory resources for different sound attributes (22–24), but otherwise place few constraints on the underlying representations.

Here, we explore whether auditory perceptual representations could be explained in part by memory limitations. One widely proposed auditory representation is that of pitch (25, 26). Pitch is

the perceptual property that enables sounds to be ordered from low to high (27), and is a salient characteristic of animal and human vocalizations, musical instrument notes, and some environmental sounds. Such sounds often contain harmonics, whose frequencies are integer multiples of a single fundamental frequency (f0) (Fig. 1A and B). Pitch is classically defined as the perceptual correlate of this f0, which is thought to be estimated from the harmonics in a sound even when the frequency component at the f0 is physically absent (Fig. 1C). Despite the prevalence of this idea in textbook accounts of pitch, there is surprisingly little direct evidence that listeners utilize representations of a sound's f0 when making pitch comparisons. For instance, discrimination of two harmonic sounds is normally envisioned to involve a comparison of estimates of the sounds' f0s (25, 28). However, if the frequencies of the sounds are altered to make them inharmonic (lacking a single f0) (Fig. 1D), discrimination remains accurate (28–31), even though there are no f0s to be compared. Such sounds do not have a pitch in the classical sense—one would not be able to consistently sing them back or otherwise match their pitch, for instance (32)—but listeners nonetheless hear a clear upward or downward change from one sound to the other, like that heard for harmonic sounds. This result is what would be expected if listeners were using the spectrum rather than the f0 (Fig. 1E), e.g., by tracking frequency shifts between sounds (33). Although harmonic advantages are evident in some other tasks plausibly related to

Significance

Sensory systems transduce information at high bandwidths but have limited memory resources, necessitating compressed storage formats. We probed auditory memory representations by measuring discrimination with and without time delays between sounds. Participants judged which of two sounds was higher. Discrimination across delays was better for harmonic sounds (containing integer multiples of a fundamental frequency, like many natural sounds) than for inharmonic sounds (with random frequencies), despite being comparably accurate without delays. Listeners appear to transition between a high-fidelity but short-lasting representation of a sound's spectrum, and a compact but enduring representation that summarizes harmonic spectra with their fundamental frequency. The results demonstrate a form of abstraction within audition whereby memory representations differ in format from representations used for on-line judgments.

Author contributions: M.J.M. and J.H.M. designed research; M.J.M. performed research; M.J.M. analyzed data; and M.J.M. and J.H.M. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

Published under the PNAS license.

¹To whom correspondence may be addressed. Email: mjmc@mit.edu or jhm@mit.edu.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2008956117/-DCSupplemental>.

First published December 1, 2020.

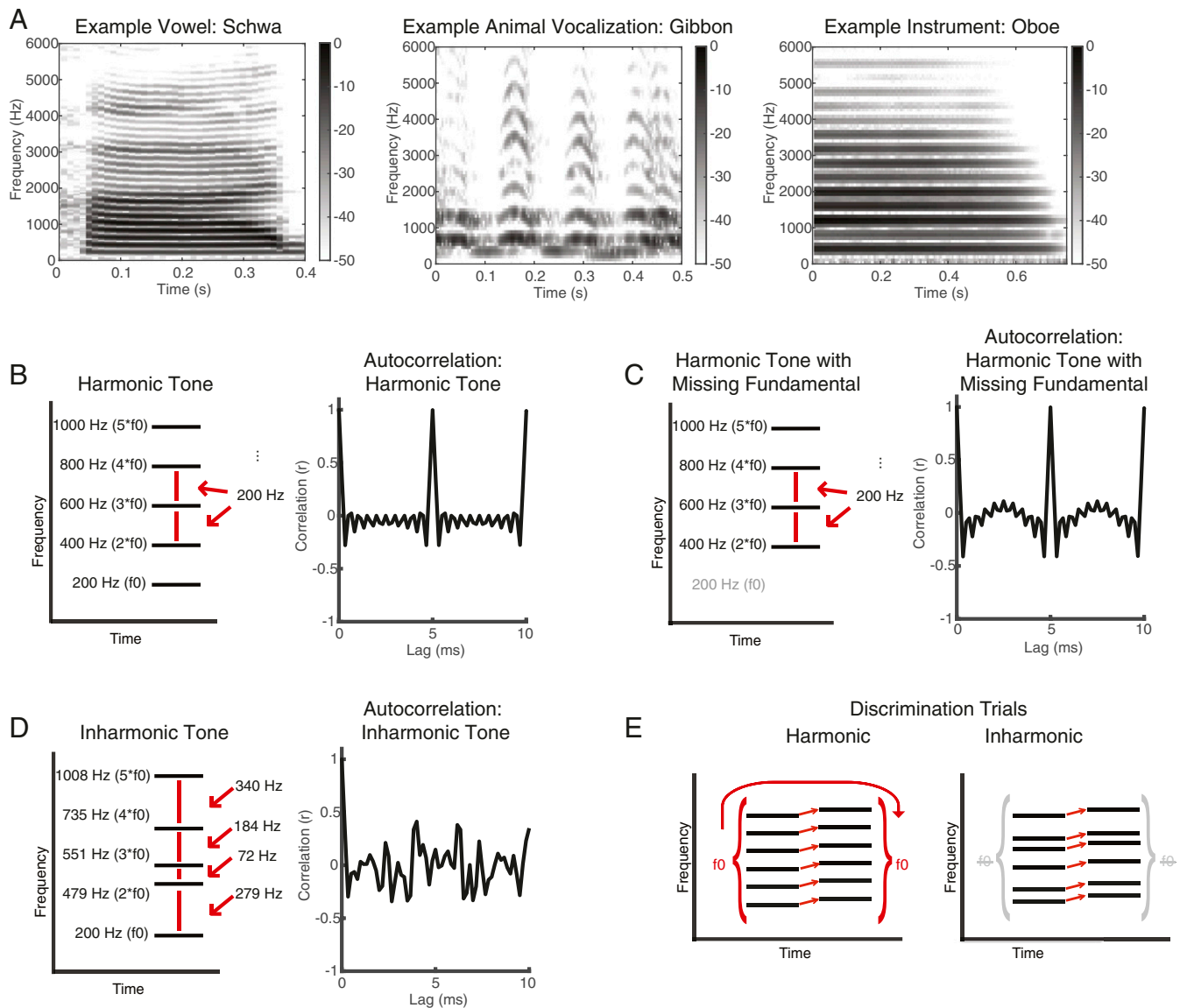


Fig. 1. Example harmonic and inharmonic sounds and discrimination trials. (A) Example spectrograms for natural harmonic sounds, including a spoken vowel, the call of a gibbon monkey, and a note played on an oboe. The components of such sounds have frequencies that are multiples of an f_0 , and as a result are regularly spaced across the spectrum. (B) Schematic spectrogram (Left) of a harmonic tone with an f_0 of 200 Hz along with its autocorrelation function (Right). The autocorrelation has a value of 1 at a time lag corresponding to the period of the tone ($1/f_0 = 5$ ms). (C) Schematic spectrogram (Left) of a harmonic tone (f_0 of 200 Hz) missing its fundamental frequency, along with its autocorrelation function (Right). The autocorrelation still has a value of 1 at a time lag of 5 ms, because the tone has the period of 200 Hz, even though this frequency is not present in its spectrum. (D) Schematic spectrogram (Left) of an inharmonic tone along with its autocorrelation function (Right). The tone was generated by perturbing the frequencies of the harmonics of 200 Hz, such that the frequencies are not integer multiples of any single f_0 in the range of audible pitch. Accordingly, the autocorrelation does not exhibit any strong peak. (E) Schematic of trials in a discrimination task in which listeners must judge which of two tones is higher. For harmonic tones, listeners could compare f_0 estimates for the two tones or follow the spectrum. The inharmonic tones cannot be summarized with f_0 s, but listeners could compare the spectra of the tones to determine which is higher.

pitch perception [such as recognizing familiar melodies, or detecting out-of-key notes (31)], the cause of this task dependence remains unclear.

In this paper, we consider whether these characteristics of pitch perception could be explained by memory constraints. One reason to estimate a sound's f_0 might be that it provides an efficient summary of the spectrum: A harmonic sound contains many frequencies, but their values can all be predicted as integer multiples of the f_0 . This summary might not be needed if two sounds are presented back-to-back, as high-fidelity (but quickly fading) sensory traces of the sounds could be compared.

However, it might become useful in situations where listeners are more dependent on a longer-lasting memory representation.

We explored this issue by measuring effects of time delay on discrimination. Our hypothesis was that time delays would cause discrimination to be based on short-term auditory memory representations (17, 34). We tested discrimination abilities with harmonic and inharmonic stimuli, varying the length of silent pauses between sounds being compared. We predicted that if listeners summarize harmonic sounds with a representation of their f_0 , then performance should be better for harmonic than for inharmonic stimuli. This prediction held across a variety of

different types of sounds and task conditions, but only when sounds were separated in time. In addition, individual differences in performance across conditions indicate that listeners switch from representing the spectrum to representing the f_0 depending on memory demands. Reliance on f_0 -based pitch thus appears to be driven in part by the need to store sound in memory.

Results

Experiment 1: Discriminating Instrument Notes with Intervening Silence. We began by measuring discrimination with and without an intervening delay between notes (Fig. 2A), using recordings of real instruments that were resynthesized to be either harmonic or inharmonic (Fig. 2B). We used real instrument sounds to maximize ecological relevance. Here and in subsequent experiments, sounds were made inharmonic by adding a random frequency “jitter” to each successive harmonic; each harmonic could be jittered in frequency by up to 50% of the original f_0 of the tone (jitter values were selected from a uniform distribution subject to constraints on the minimum spacing between adjacent frequencies). Making sounds inharmonic renders them inconsistent with any f_0 , such that the frequencies cannot

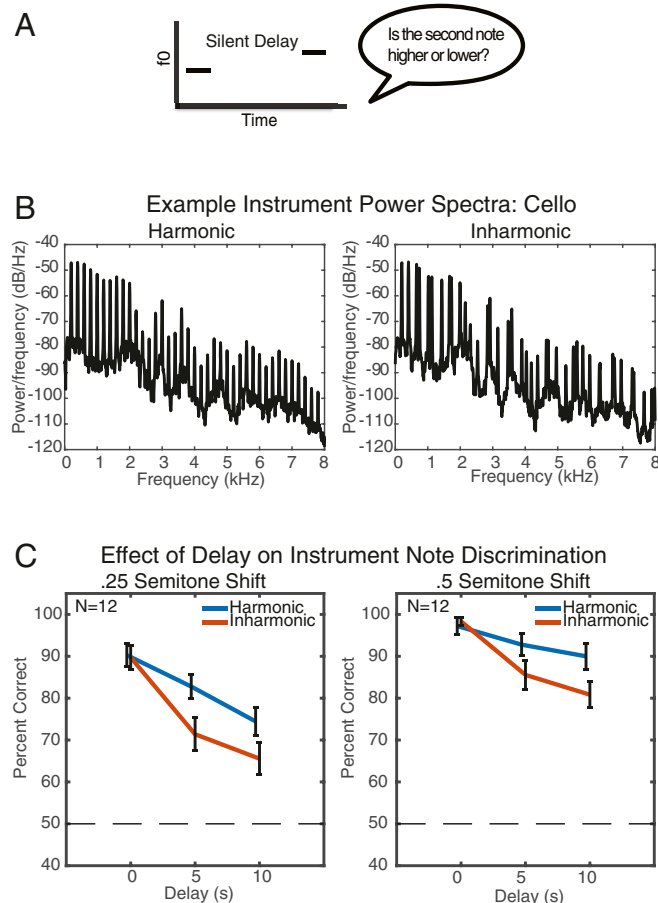


Fig. 2. Experiment 1: Harmonic advantage when discriminating instrument notes across a delay. (A) Schematic of trial structure for experiment 1. During each trial, participants heard two notes played by the same instrument and judged whether the second note was higher or lower than the first note. Notes were separated by a delay of 0, 5, or 10 s. (B) Power spectra of example harmonic and inharmonic (with frequencies jittered) notes from a cello (the fundamental frequency of the harmonic note is 200 Hz in this example). (C) Results of experiment 1 plotted separately for the two difficulty levels that were used. Error bars show SEM.

be summarized by a single f_0 . Whereas the autocorrelation function of a harmonic tone shows a strong peak at the period of the f_0 (Fig. 1B and C), that for an inharmonic tone does not (Fig. 1D). The same pattern of random jitter was added to each of the two notes in a trial. There was thus a direct correspondence between the frequencies of the first and second notes even though the inharmonic stimuli lacked an f_0 (Fig. 1E).

Participants heard two notes played by the same instrument (randomly selected on each trial from the set of cello, baritone saxophone, ukulele, pipe organ, and oboe, with the instruments each appearing an equal number of times within a given condition). The two notes were separated by 0, 5, or 10 s of silence. Notes always differed by either a quarter of a semitone ($\sim 1.5\%$ difference between the note f_0 s) or a half semitone ($\sim 3\%$ difference between the note f_0 s). Participants judged whether the second note was higher or lower than the first.

We found modest decreases in performance for harmonic stimuli as the delay increased (significant main effect of delay for both 0.25 semitone [$F_{(2,22)} = 25.19, P < 0.001, \eta_p^2 = 0.70$, Fig. 2C] and 0.5 semitone conditions [$F_{(2,22)} = 5.31, P = 0.01, \eta_p^2 = 0.33$]). This decrease in performance is consistent with previous studies examining memory for complex tones (21, 35). We observed a more pronounced decrease in performance for inharmonic stimuli, with worse performance than for harmonic stimuli at both delays [5 s: $t_{(11)} = 4.48, P < 0.001$ for 0.25 semitone trials, $t_{(11)} = 2.12, P = 0.057$ for 0.5 semitone trials; 10 s: $t_{(11)} = 3.64, P = 0.004$ for 0.25 semitone trials, $t_{(11)} = 4.07, P = 0.002$ for 0.5 semitone trials] despite indistinguishable performance for harmonic and inharmonic sounds without a delay [$t_{(11)} = 0.43, P = 0.67$ for 0.25 semitone trials, $t_{(11)} = -0.77, P = 0.46$ for 0.5 semitone trials]. These differences produced a significant interaction between the effect of delay and that of harmonicity [$F_{(2,22)} = 7.66, P = 0.003, \eta_p^2 = 0.41$ for 0.25 semitone trials; $F_{(2,22)} = 3.77, P = 0.04, \eta_p^2 = 0.26$ for 0.5 semitone trials]. This effect was similar for musicians and nonmusicians (SI Appendix, Fig. S1). Averaging across the two difficulty conditions, we found no main effect of musicianship [$F_{(1,10)} = 3.73, P = 0.08, \eta_p^2 = 0.27$] and no interaction between musicianship, harmonicity, and delay length [$F_{(2,20)} = 0.58, P = 0.57, \eta_p^2 = 0.06$]. Moreover, the interaction between delay and harmonicity was significant in nonmusicians alone [$F_{(2,10)} = 6.48, P = 0.02, \eta_p^2 = 0.56$]. This result suggests that the harmonic advantage is not dependent on extensive musical training. Overall, the results of experiment 1 are consistent with the idea that a sound’s spectrum can mediate discrimination over short time intervals (potentially via a sensory trace of the spectrum), but that memory over longer periods relies on a representation of f_0 .

Experiment 2: Discriminating Synthetic Tones with Intervening Silence. We replicated and extended the results of experiment 1 using synthetic tones, the acoustic features of which can be more precisely manipulated. We generated complex tones that were either harmonic or inharmonic, applying fixed bandpass filters to all tones in order to minimize changes in the center of mass of the tones that could otherwise be used to perform the task (Fig. 3A). The first audible harmonic of these tones was generally the fourth (although it could be the third or fifth, depending on the f_0 or jitter pattern). To gauge the robustness of the effect across difficulty levels, we again used two f_0 differences (0.25 and 0.5 semitones).

To further probe the effect of delay on representations of f_0 , we included a third condition (“Interleaved Harmonic”) where each of the two tones on a trial contained half of the harmonic series (Fig. 3A). One tone always contained harmonics [1, 4, 5, 8, 9, etc.], and the other always contained harmonics [2, 3, 6, 7, 10, etc.] (28). The order of the two sets of harmonics was randomized across trials. This selection of harmonics eliminates common harmonics between tones. While in the Harmonic and

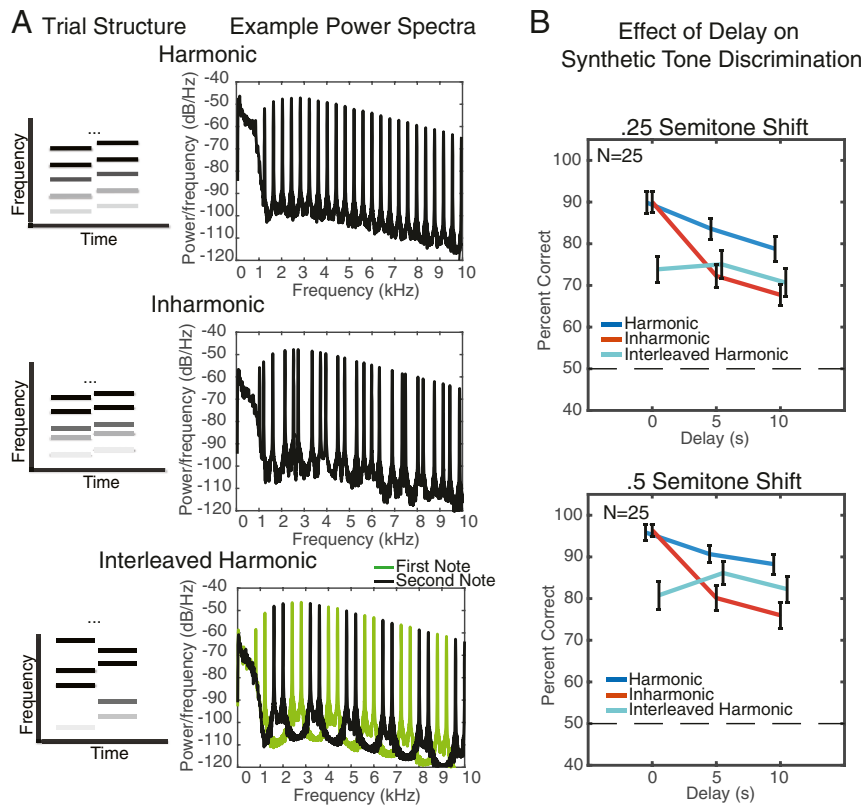


Fig. 3. Experiment 2: Harmonic advantage when discriminating synthetic tones across a delay. (A) Schematic of stimuli and trial structure in experiment 2. (Left) During each trial, participants heard two tones and judged whether the second was higher or lower than the first. Tones were separated by a delay of 0, 5, or 10 s. (Right) Power spectra of 400-Hz tones for Harmonic, Inharmonic, and Interleaved Harmonic conditions. (B) Results of experiment 2, plotted separately for the two difficulty levels. Error bars show SEM.

Inharmonic conditions listeners could use the correspondence of the individual harmonics to compare the tones, in the Interleaved Harmonic condition there is no direct correspondence between harmonics present in the first and second notes, such that the task can only be performed by estimating and comparing the f0s of the tones. By applying the same bandpass filter to each Interleaved Harmonic tone, we sought to minimize timbral differences that are known to impair f0 discrimination (36), although the timbre nonetheless changed somewhat from note-to-note, which one might expect would impair performance to some extent. Masking noise was included in all conditions to prevent distortion products from being audible, which might otherwise be used to perform the task. The combination of the bandpass filter and the masking noise was also sufficient to prevent the frequency component at the f0 from being used to perform the task.

As in experiment 1, harmonic and inharmonic tones were similarly discriminable with no delay (Fig. 3B; $Z = 0.55$, $P = 0.58$ for 0.25 semitone trials; $Z = 0.66$, $P = 0.51$ for 0.5 semitone trials, Wilcoxon signed-rank test). However, performance with inharmonic tones was again significantly worse than performance with harmonic tones with a delay (5-s delay: $Z = 3.49$, $P < 0.001$ for 0.25 semitones, $Z = 3.71$, $P < 0.001$ for 0.5 semitones; 10-s delay: $Z = 3.46$, $P < 0.001$ for 0.25 semitones, $Z = 4.01$, $P < 0.001$ for 0.5 semitones), yielding interactions between the effect of delay and harmonicity [$F_{(2,48)} = 10.71$, $P < 0.001$, $\eta_p^2 = 0.31$ for 0.25 semitones; $F_{(2,48)} = 26.17$, $P < 0.001$, $\eta_p^2 = 0.52$ for 0.5 semitones; P values calculated via bootstrap because data were nonnormal].

Performance without a delay was worse for interleaved-harmonic tones than for the regular harmonic and inharmonic

tones, but unlike in those other two conditions, Interleaved Harmonic performance did not deteriorate significantly over time (Fig. 3B). There was no main effect of delay for either 0.25 semitones [$F_{(2,48)} = 1.23$, $P = 0.30$, $\eta_p^2 = 0.05$] or 0.5 semitones [$F_{(2,48)} = 2.56$, $P = 0.09$, $\eta_p^2 = 0.10$], in contrast to the significant main effects of delay for Harmonic and Inharmonic conditions at both difficulty levels ($P < 0.001$ in all cases). This pattern of results is consistent with the idea that there are two representations that listeners could use for discrimination: a sensory trace of the spectrum, which decays quickly over time, and a representation of the f0, which is better retained over a delay. The spectrum can be used in the Harmonic and Inharmonic conditions, but not in the Interleaved Harmonic condition.

As in experiment 1, the effects were qualitatively similar for musicians and nonmusicians (SI Appendix, Fig. S2). Although there was a significant main effect of musicianship [$F_{(1,23)} = 10.28$, $P < 0.001$, $\eta_p^2 = 0.99$], the interaction between the effects of delay and harmonicity was significant in both musicians [$F_{(2,28)} = 20.44$, $P < 0.001$, $\eta_p^2 = 0.59$] and nonmusicians [$F_{(2,18)} = 11.99$, $P < 0.001$, $\eta_p^2 = 0.57$], and there was no interaction between musicianship, stimulus type (Harmonic, Inharmonic, Interleaved Harmonic), and delay length [$F_{(4,92)} = 0.19$, $P = 0.98$, $\eta_p^2 = 0.01$].

Experiment 3: Discriminating Synthetic Tones with a Consistent Inharmonic Spectrum. The purpose of experiment 3 was to examine the effects of the specific inharmonicity manipulation used in experiments 1 and 2. In experiments 1 and 2, we used a different inharmonic jitter pattern for each trial. In principle, listeners might be able to learn a spectral pattern if it repeats

across trials (37, 38), such that the harmonic advantage found in experiments 1 and 2 might not have been due to harmonicity per se, but rather to the fact that the harmonic pattern occurred repeatedly whereas the inharmonic pattern did not. In experiment 3, we compared performance in a condition where the jitter pattern was held constant across trials (“Inharmonic-Fixed”) vs. when it was altered every trial (although the same for the two notes of a trial), as in the previous experiments (“Inharmonic”).

Unlike the first two experiments, we used adaptive procedures to measure discrimination thresholds. Adaptive threshold measurements avoided difficulties associated with choosing stimulus differences appropriate for the anticipated range of performance across the conditions. Participants completed 3-down-1-up two-alternative-forced-choice (“is the second tone higher or lower than the first”) adaptive “runs,” each of which produced a single threshold measurement. Tones were either separated by no delay, a 1-s delay, or a 3-s delay (Fig. 4A). There were three stimulus conditions, separated into three blocks: Harmonic (as in experiment 2); Inharmonic, where the jitter changed for every trial (as in experiment 2); and Inharmonic-Fixed, where a single random jitter pattern (chosen independently for each participant) was used across the entire block of adaptive runs. Participants completed four adaptive runs for each delay and stimulus pair. Delay conditions were intermixed within the Harmonic, Inharmonic, and Inharmonic-Fixed stimulus condition blocks, resulting in 12 adaptive runs per block (Fig. 4B), with ~60 trials per run on average. The order of these three blocks was randomized across participants. This design does not preclude the possibility that participants might learn a repeating inharmonic pattern given even more exposure to it, but it puts the inharmonic and harmonic tones on equal footing, testing whether the harmonic advantage might be due to relatively short-term learning of the consistent spectral pattern provided by harmonic tones.

As shown in Fig. 4C, thresholds were similar for all conditions with no delay (no significant difference between any of the

conditions; $Z < 0.94$, $P > 0.34$ for all pairwise comparisons, Wilcoxon signed-rank test, used because the distribution of thresholds was nonnormal). These discrimination thresholds were comparable to previous measurements for stimuli with resolved harmonics (39, 40). However, thresholds were slightly elevated (worse) for both the Inharmonic and Inharmonic-Fixed conditions with a 1-s delay (Inharmonic: $Z = 2.40$, $P = 0.016$; Inharmonic-Fixed: $Z = 1.92$, $P = 0.055$), and were much worse in both conditions with a 3-s delay (Inharmonic: $Z = 3.41$, $P < 0.001$; Inharmonic-Fixed: $Z = 2.71$, $P = 0.007$). There was no significant effect of the type of inharmonicity [$F_{(1,22)} = 0.94$, $P = 0.34$, $\eta_p^2 = 0.04$, comparing Inharmonic vs. Inharmonic-Fixed conditions], providing no evidence that participants learn to use specific jitter patterns over the course of a typical experiment duration. We again observed significant interactions between the effects of delay and stimulus type (Harmonic, Inharmonic, and Inharmonic-Fixed conditions) in both musicians [$F_{(4,44)} = 3.85$, $P = 0.009$, $\eta_p^2 = 0.26$] and nonmusicians [$F_{(4,40)} = 3.04$, $P = 0.028$, $\eta_p^2 = 0.23$], and no interaction between musicianship, stimulus type, and delay length [$F_{(4,84)} = 1.05$, $P = 0.07$, $\eta_p^2 = 0.05$; *SI Appendix*, Fig. S3].

These results indicate that the harmonic advantage cannot be explained by the consistency of the harmonic spectral pattern across trials, as making the inharmonic spectral pattern consistent did not reduce the effect. Given this result, in subsequent experiments we opted to use Inharmonic rather than Inharmonic-Fixed stimuli, to avoid the possibility that the results might otherwise be biased by the choice of a particular jitter pattern.

Experiment 4: Discriminating Synthetic Tones with a Longer Intertrial Interval. To assess whether the inharmonic deficit could somehow reflect interference from successive trials (41) rather than the decay of memory during the interstimulus delay, we replicated a subset of the conditions from experiment 3 using a longer intertrial interval. We included only the Harmonic and Inharmonic conditions, with and without a 3-s delay between notes. For each

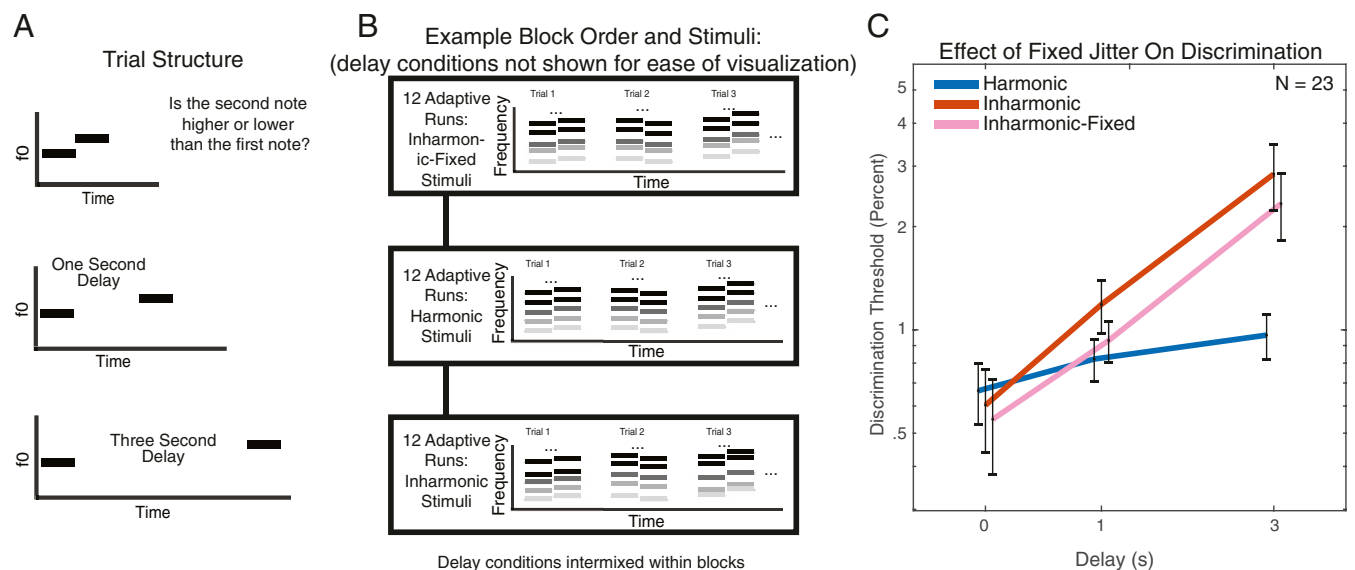


Fig. 4. Experiment 3: Harmonic advantage persists for consistent inharmonic jitter pattern. (A) Schematic of trial structure for experiment 3. Task was identical to that of experiments 1 and 2, but with delay durations of 0, 1, and 3 s, and adaptive threshold measurements rather than method of constant stimuli. (B) Example block order for experiment 3, in which the 12 adaptive runs for each condition were presented within a contiguous block. The beginning of an example run is shown schematically for each type of condition. Stimulus conditions (Harmonic, Inharmonic, Inharmonic-Fixed) were blocked. Note the difference between the Inharmonic-Fixed condition, in which the same jitter pattern was used across all trials within a block, and the Inharmonic condition, in which the jitter pattern was different on every trial. Delay conditions were intermixed within each block. (C) Results of experiment 3. Error bars show within-subject SEM.

condition, participants completed four adaptive threshold measurements without any enforced delay between trials, and four adaptive measurements where we imposed a 4-s intertrial interval (such that the delay between trials would always be at least 1 s longer than the delay between notes of the same trial). This experiment was run online because the laboratory was temporarily closed due to the COVID-19 virus.

The interaction between within-trial delay (0 vs. 3 s) and stimulus type (Harmonic vs. Inharmonic) was present both with and without the longer intertrial interval [with: $F_{(1,37)} = 4.92, P = 0.03, \eta_p^2 = 0.12$; without: $F_{(1,37)} = 12.34, P = 0.001, \eta_p^2 = 0.25$]. In addition, we found no significant interaction between intertrial interval, within-trial delay, and harmonic vs. inharmonic stimuli [$F_{(1,37)} = 1.78, P = 0.19, \eta_p^2 = 0.05$; *SI Appendix, Fig. S4*]. This result suggests that the inharmonic deficit is due to difficulties retaining a representation of the tones during the delay period, rather than some sort of interference from preceding stimuli.

Experiment 5: One-Shot Discrimination with a Longer Intervening Delay. Experiments 1 to 4 leave open the possibility that listeners might use active rehearsal of the stimuli (singing to themselves, for instance) to perform the task over delays. Although prior results suggest that active rehearsal does not obviously aid discrimination of tones over delays (17, 42), it could in principle explain the harmonic advantage on the assumption that it is more difficult to rehearse an inharmonic stimulus, and so it seemed important to address. To assess whether the harmonic advantage reflects rehearsal, we ran a “one-shot” online experiment with much longer delay times, during which participants filled out a demographic survey. We assumed this unrelated task would prevent them from actively rehearsing the heard tone. This experiment was run online to recruit the large number of participants needed to obtain sufficient power. We have previously found that online participants can perform about as well as in-laboratory participants (38, 43) provided basic steps are taken both to maximize the chances of reasonable sound presentation by testing for earphone/headphone use (44), and to ensure compliance with instructions, either by providing training or by removing poorly performing participants using hypothesis-neutral screening procedures.

Each participant completed only two trials in the main experiment. One trial had no delay between notes, as in the 0-s delay conditions of the previous experiments. During the other trial participants heard one tone, then were redirected to a short demographic survey, and then heard the second tone (Fig. 5A). The order of the two trials was randomized across participants. For each participant, both trials contained the same type of tone, randomly assigned. The tones were either harmonic, inharmonic, or interleaved-harmonic (each identical to the tones used in experiment 2). The two test tones always differed in f_0 by a semitone. The discrimination task was described to participants at the start of the experiment, such that participants knew they should try to remember the first tone before the survey and that they would be asked to compare it to a second tone heard after the survey. To ensure task comprehension, participants completed 10 practice trials with feedback (without a delay, with an f_0 difference of a semitone). These practice trials were always with same types of tones the participant would hear during the main experiment (for instance, if participants heard inharmonic stimuli in the test trials, the practice trials also featured inharmonic stimuli).

We ran a large number of participants to obtain sufficient power given the small number of trials per participant. Participants completed the survey at their own pace. We measured the time interval between the onset of the first note before the survey and the onset of the second note after the survey. Prior to analysis, we removed participants who completed the survey in under 20 s (proceeding through the survey so rapidly as to

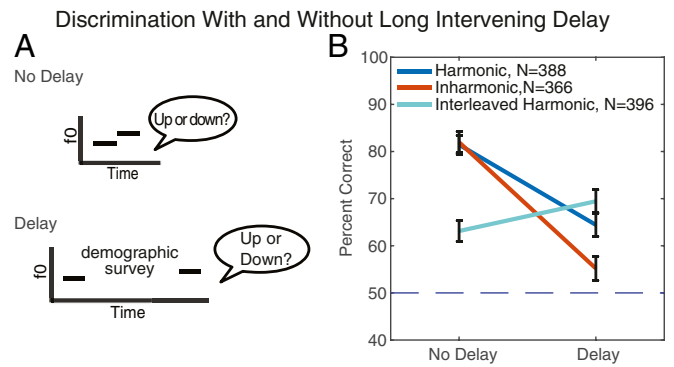


Fig. 5. Experiment 5: Harmonic advantage persists over longer delays with intervening task. (A) Schematic of the two trial types in experiment 5. Task was identical to that of experiment 2. However, for the “delay” condition participants were redirected to a short demographic survey that they could complete at their own pace. (B) Results of experiment 5 (which had the same stimulus conditions as experiment 2). Error bars show SEM, calculated via bootstrap.

suggest that the participant did not read the questions; 1 participant), or participants who took longer than 3 min (17 participants). Of the remaining 1,150 participants, the mean time spent on the survey was 58.9 s (SD of 25.1, median of 52.8 s, and median absolute deviation of 18.4 s). There was no significant difference between the time taken for each of the conditions ($Z \leq 0.94, P \geq 0.345$ for all pairwise comparisons).

As shown in Fig. 5B, experiment 5 qualitatively replicated the results of experiments 1 and 2 even with the longer delay period and concurrent demographic survey. Without a delay, there was no difference between performance with harmonic and inharmonic conditions ($P = 0.61$, via bootstrap). With a delay, performance in both the Harmonic and Inharmonic conditions was worse than without a delay ($P < 0.0001$ for both). However, this impairment was larger for the Inharmonic condition; performance with inharmonic tones across a delay was significantly worse than that with harmonic tones ($P < 0.001$), producing an interaction between the type of tone and delay ($P = 0.02$). By contrast, performance on the Interleaved Harmonic condition did not deteriorate over the delay and in fact slightly improved ($P = 0.004$). This latter result could reflect the decay of the representation of the spectrum of the first tone across the delay, which in this condition might otherwise impair f_0 discrimination by providing a competing cue [because the spectra of the two tones are different (28)].

Experiment 6: Individual Differences in Tone Discrimination. The similarity in performance between harmonic and inharmonic tones without a delay provides circumstantial evidence that listeners are computing changes in the same way for both types of stimuli, presumably using a representation of the spectrum in both cases. However, the results leave open the alternative possibility that listeners use a representation of the f_0 for the harmonic tones despite having access to a representation of the spectrum (which they must use with the inharmonic tones), with the two strategies happening to support similar accuracy.

To address these possibilities, and to explore whether listeners use different encoding strategies depending on memory constraints, we employed an individual differences approach (45, 46). The underlying logic is that performance on tasks that rely on the same perceptual representations should be correlated across participants. For example, if two discrimination tasks rely on similar representations, a participant with a low threshold on one task should tend to have a low threshold on the other task.

In experiment 6, we estimated participants' discrimination thresholds either with or without a 3-s delay between stimuli, using a 3-down-1-up adaptive procedure. Stimuli included harmonic, inharmonic, and interleaved-harmonic complex tones, generated as in experiments 2 to 5 (Fig. 6A). We used inharmonic stimuli for which a different pattern of jitter was chosen for each trial, because experiment 3 showed similar results whether the inharmonic pattern was consistent or not, and because it seemed better to avoid a consistent jitter. Specifically, it seemed possible that fixing the jitter pattern across trials for each participant might create artifactual individual differences given that some jitter patterns are by chance closer to harmonic than others. Having the jitter pattern change from trial to trial produces a similar distribution of stimuli across participants and should average out the effects of idiosyncratic jitter patterns. We also included a condition with pure tones (sinusoidal tones, containing a single frequency) at the frequency of the fourth harmonic of the tones in the Harmonic condition, and a fifth condition where each note contained two randomly chosen harmonics from each successive set of four (1 to 4, 5 to 8, etc.). By chance, some harmonics could be found in both notes. This condition (Random Harmonic) was intended as an internal replication of the anticipated results with the Interleaved Harmonic condition.

The main hypothesis we sought to test was that listeners use one of two different representations depending on the duration of the interstimulus delay and the nature of the stimulus. Specifically, without a delay, listeners use a detailed spectral representation for both harmonic and inharmonic sounds, relying on an f0-based representation only when a detailed spectral pattern is not informative (as in the Interleaved Harmonic condition). In the presence of a delay, they switch to relying on an f0-based representation for all harmonic sounds.

The key predictions of this hypothesis in terms of correlations between thresholds are 1) high correlations between Harmonic and Inharmonic discrimination without a delay, 2) lower correlations between Harmonic and Inharmonic discrimination with a delay, 3) low correlations between Interleaved Harmonic discrimination and both the Harmonic and Inharmonic conditions with no delay (because the former requires a representation of the f0), and 4) higher correlations between Interleaved Harmonic and Harmonic discrimination with a delay.

We ran this study online to recruit sufficient numbers to measure the correlations of interest. This experiment was relatively arduous (it took ~2 h to complete), and based on pilot experiments we anticipated that many online participants would perform poorly relative to in-laboratory participants, perhaps because the chance of distraction occurring at some point over the 2 h is high. To obtain a criterion level of performance with which to determine inclusion for online participants, we ran a group of participants in the laboratory to establish acceptable performance levels. We calculated the overall mean threshold for all conditions without a delay (five conditions) for the best two-thirds of in-laboratory participants, and excluded online participants whose mean threshold on the first run of those same five conditions (without a delay) was above this mean in-laboratory threshold. To avoid double dipping, we subsequently analyzed only the last three threshold measurements for each condition from online participants (10 total conditions, three runs for each condition). This inclusion procedure selected 164 of 450 participants for the final analyses shown in Fig. 6B–E. See *SI Appendix, Fig. S5* for results with a less stringent inclusion criterion (the critical effects remained present with this less stringent criterion).

Mean thresholds. The pattern of mean threshold measurements obtained online qualitatively replicated the results of experiments 1 to 5 (Fig. 6B). Inharmonic thresholds were indistinguishable from Harmonic thresholds without a delay ($Z = 1.74$,

$P = 0.08$), but were higher when there was a delay between sounds ($Z = -9.07$, $P < 0.001$). This produced a significant interaction between effects of tone type and delay [$F_{(1,163)} = 88.45$, $P < 0.001$, $\eta_p^2 = 0.35$]. In addition, as in the previous experiments, there was no significant effect of delay for the Interleaved Harmonic condition (Interleaved Harmonic condition with vs. without delay; $Z = 0.35$, $P = 0.72$).

Individual differences—harmonic, inharmonic, and interleaved harmonic conditions. Fig. 6C shows the correlations across participants between different pairs of thresholds. Thresholds were correlated to some extent for all pairs of conditions, presumably reflecting general factors such as attention or motivation that produce variation in performance across participants. However, some correlations were higher than others. Fig. 6D and E plots the correlations (extracted from the matrix in Fig. 6C) for which our hypothesis makes critical predictions, to facilitate their inspection. Correlations here and elsewhere were corrected for the reliability of the underlying threshold measurements (47); the correlation between two thresholds was divided by the square root of the product of their reliabilities (Cronbach's α calculated from Spearman's correlations between pairs of the last three runs of each condition). This denominator provides a ceiling for each correlation, as the correlation between two variables is limited by the accuracy with which each variable is measured. Thus, correlations could in principle be as high as 1 in the limit of large data, but because the threshold reliabilities were calculated from modest sample sizes, the corrected correlations could in practice slightly exceed 1.

Performance on Harmonic and Inharmonic conditions without a delay was highly correlated across participants ($\rho = 0.88$, $P < 0.001$). However, the correlation between Harmonic and Inharmonic conditions with a delay was substantially lower ($\rho = 0.73$, $P < 0.001$, significant difference between Harmonic–Inharmonic correlations with and without a delay, $P = 0.007$, calculated via bootstrap). We observed the opposite pattern for the Harmonic and the Interleaved Harmonic conditions: the correlation between Harmonic and Interleaved Harmonic thresholds was lower without a delay ($\rho = 0.74$, $P < 0.001$) than with a delay ($\rho = 0.86$, $P < 0.001$, significant difference between correlations, $P = 0.03$). This pattern of results yielded a significant interaction between the conditions being compared and the effect of delay (difference of differences between correlations with and without a delay = 0.27, $P = 0.019$). We replicated this interaction in a pilot version of the experiment that featured slightly different stimulus conditions and intervening notes in the delay period (*SI Appendix, Fig. S6*; $P = 0.006$). Overall, the results of the individual differences analysis indicate that when listening to normal harmonic tones, participants switch between two different pitch mechanisms depending on the time delay between tones.

Results with pure tones and random harmonic condition. The results for Pure Tones were similar to those for the Harmonic condition (Fig. 6E), with quantitatively indistinguishable mean thresholds (without delay: $Z = 0.16$, $P = 0.87$; with delay: $Z = 1.18$, $P = 0.24$). Moreover, the correlations between Harmonic and Pure Tone thresholds were high both with and without a delay, and not significantly different ($P = 0.13$). These results suggest that similar representations are used to discriminate pure and harmonic complex tones. In addition, the correlations between the Pure Tone condition and the Inharmonic and Interleaved Harmonic conditions showed a similar pattern to that for the Harmonic condition (compare Fig. 6D and E), producing a significant interaction between the conditions being compared and the effect of delay (difference of differences between correlations with and without a delay = 0.32, $P = 0.006$). This interaction again replicated in the pilot version of the experiment (*SI Appendix, Fig. S6*; $P = 0.008$).

The Random Harmonic results largely replicated the findings from the Interleaved Harmonic condition (Fig. 6C). Evidently,

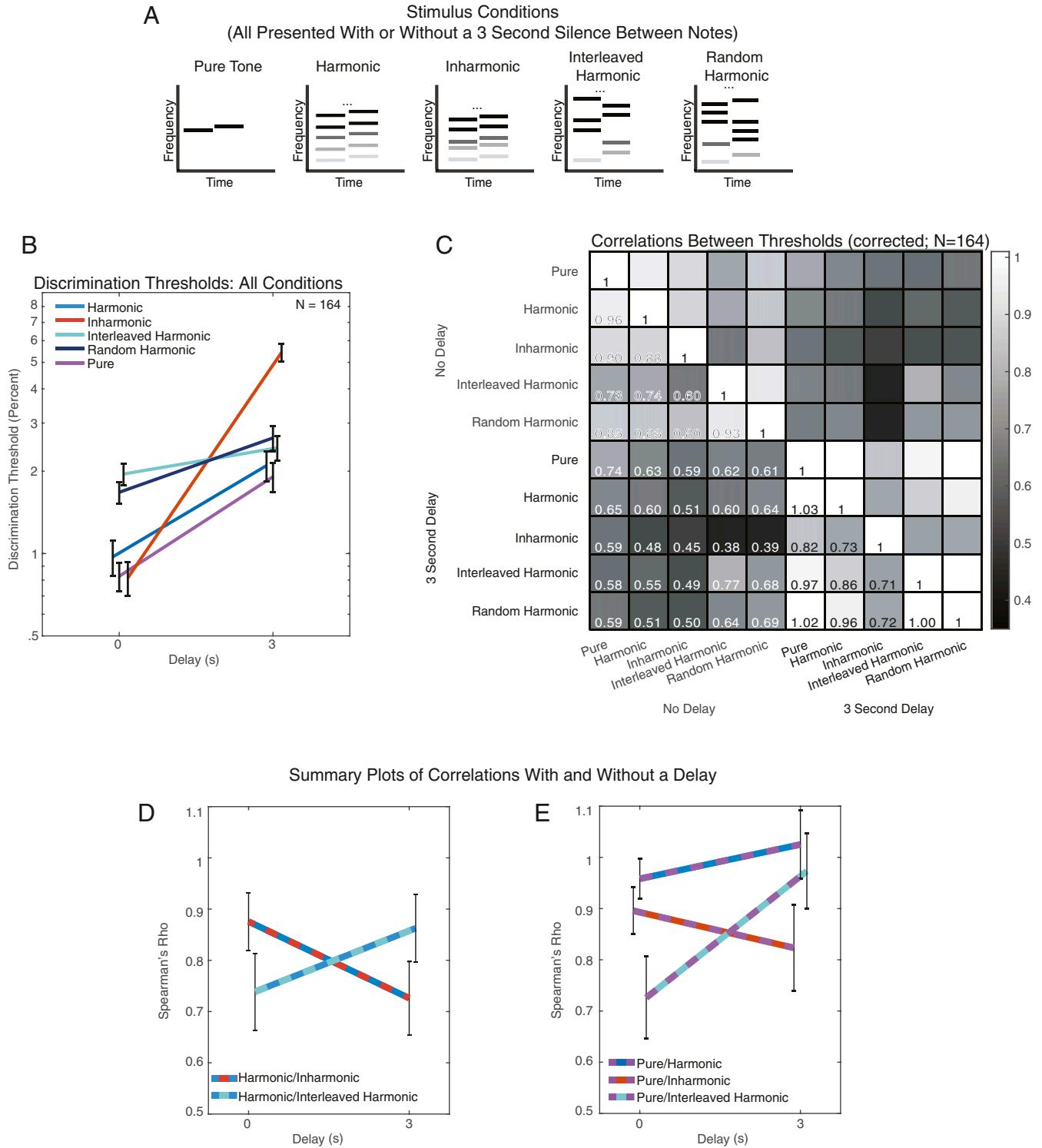


Fig. 6. Experiment 6: Individual differences suggest different representations depending on memory demands. (A) Schematic of the five types of stimuli used in experiment 6. Task and procedure (adaptive threshold measurements) were identical to those of experiment 4, but with no intertrial delays. (B) Discrimination thresholds for all stimulus conditions, with and without a 3-s delay between tones. Here, and in D and E, error bars show SEM, calculated via bootstrap. (C) Matrix of the correlations between thresholds for all pairs of conditions. Correlations are Spearman's ρ , corrected for the reliability of the threshold measurements (i.e., corrected for attenuation). Corrected correlations can slightly exceed 1 given that they are corrected with imperfect estimates of the reliabilities. (D) Comparison between Harmonic/Inharmonic and Harmonic/Interleaved Harmonic threshold correlations, with and without a delay. (E) Comparison between Pure Tone condition correlations with and without a delay.

the changes that we introduced in the harmonic composition of the tones being compared in this condition were sufficient to preclude the use of the spectrum, and, as in the Interleaved Harmonic condition, performance was determined by f0-based pitch regardless of the delay between notes.

Discussion

We examined the relationship between pitch perception and memory by measuring the discrimination of harmonic and inharmonic sounds with and without a time delay between stimuli. Across several experiments, we found that discrimination over a delay was better for harmonic sounds than for inharmonic sounds, despite comparable accuracy without a delay. This effect was observed over delays of a few seconds and persisted over longer delays with an intervening distractor task. We also analyzed individual differences in discrimination thresholds across a large number of participants. Harmonic and inharmonic discrimination thresholds were highly correlated without a delay between sounds but were less correlated with a delay. By contrast, thresholds for harmonic tones and tones designed to isolate f0-based discrimination (interleaved-harmonic tones) showed the opposite pattern, becoming more correlated with a delay between sounds. Together, the results suggest that listeners use different representations depending on memory demands, comparing spectra for sounds nearby in time, and f0s for sounds separated in time. The results provide evidence for two distinct mechanisms for pitch discrimination, reveal the constraints that determine when they are used, and demonstrate a form of abstraction within the auditory system whereby the representations of memory differ in format from those used for rapid on-line judgments about sounds.

In hearing research, the word “pitch” has traditionally referred to the perceptual correlate of the f0 (26). In some circumstances, listeners must base behavior on the absolute f0 of a sound of interest, as when singing back a heard musical note. However, much of the time the information that matters to us is conveyed by how the f0 changes over time, and our results indicate that listeners often extract this information using a representation of the spectrum rather than the f0. One consequence of this is that note-to-note changes can be completely unambiguous even for inharmonic sounds that lack an unambiguous f0. Is this pitch perception? Under typical listening conditions (where sounds are harmonic), the changes in the spectrum convey changes in f0, and thus enable judgments about how the f0 changes from note to note. Consistent with this idea, listeners readily describe what they hear in the inharmonic conditions of our experiments as a pitch change, as though the consistent shift in the spectrum is interpreted as a change in the f0 even though neither note has a clear f0. We propose that these spectral judgments should be considered part of pitch perception, which we construe to be the set of computations that enable judgments about a sound’s f0. The perception of inharmonic “pitch changes” might thus be considered an illusion, exploiting the spectral pitch mechanism in conditions in which it does not normally operate.

Why do listeners not base pitch judgments of harmonic sounds on their f0s when sounds are back-to-back? One possibility is that representations of f0 are in some cases less accurate and produce poorer discrimination than those of the spectrum. The results with interleaved harmonics (stimuli designed to isolate f0-based pitch) are consistent with this idea, as discrimination without a delay was worse for interleaved harmonics than for either harmonic or inharmonic tones that had similar spectral composition across notes. However, we note that this deficit could also reflect other stimulus differences, such as the potentially interfering effects of the changes in the spectrum from note to note (36). Regardless of the root cause for the reliance on spectral representations, the fact that performance with interleaved harmonics was similar with and without modest delays

suggests that representations of the f0 are initially available in parallel with representations of the spectrum, with task demands determining which is used in the service of behavior. As time passes, the high-fidelity representation of the spectrum appears to degrade, and listeners switch to more exclusively using a representation of the f0.

Relation to Previous Studies of Pitch and Pitch Memory. The use of time delays to study memory for frequency and/or pitch has a long tradition (48, 49). Our results here are broadly consistent with this previous work, but provide evidence for differences in how representations of the f0 and the spectrum are retained over time. A number of studies have examined memory for tones separated by various types of interfering stimuli, and collectively provide evidence that the f0 of a sound is retained in memory. For example, intervening sounds interfere most with discrimination if their f0s are similar to the tones being discriminated, irrespective of whether the intervening sounds are speech or synthetic tones (22), and irrespective of the spectral content of the intervening notes or the two comparison tones (21). Our results complement these findings by showing that memory for f0 has different characteristics from that for frequency content (spectra), by showing how these differences impact pitch perception, and by suggesting that memory for f0 should be viewed as a form of compression. We also show that introducing a delay between notes forces listeners to use the f0 rather than the spectrum, which may be useful in experimentally isolating f0-based pitch in future studies.

Our results are consistent with the idea that memory capacity limitations for complex spectra in some cases limit judgments about sound. Previous studies of memory for complex tones failed to find clear evidence for such capacity limitations, in that there was no interaction between the effects of interstimulus interval and of the number of constituent frequencies on the accuracy of judgments of a remembered tone (18). However, there were many differences between these prior experiments and those described here that might explain the apparent discrepancy, including that the time intervals tested were short (at most 2 s) compared to those in our experiments, and the participants highly practiced. It is possible that, under such conditions, listeners are less dependent on the memory representations that were apparently tapped in our experiments. The tasks used in those prior experiments were also different (involving judgments of a single frequency component within an inharmonic complex tone), as were the stimuli (frequencies equidistant on a logarithmic scale). Quantitative models of memory representations and their use in behavioral tasks seem likely to be an important next step in evaluating whether the available results can be explained by a single type of memory store.

Relation to Visual Memory. Our results could have interesting analogs in vision. Visual short-term memory has been argued to store relatively abstract representations (1–4, 6, 7), and the grouping of features into object-like representations is believed to increase its effective capacity (50–52). Our results raise the question of whether such benefits are specific to memory. It is plausible that for stimuli presented back-to-back, discrimination of visual element arrays would be similar irrespective of the element arrangement, with advantages for elements that are grouped into a coherent pattern only appearing when short-term memory is taxed. To our knowledge, this has not been explicitly tested.

One apparent difference between auditory and visual memory is that in some contexts visual memory for simple stimuli decays relatively slowly, with performance largely unimpaired for multisecond delays comparable to those used here (24). By contrast, auditory memory is more vulnerable, with performance decreases often evident over seconds even for pure tone discrimination (e.g., Fig. 6B). As a consequence, visual memory has often

been studied via memory “masking” effects in which stimuli presented in the delay period impair performance if they are sufficiently similar to the stimuli being remembered (53–55). Such effects also occur for auditory memory (20–22), but the performance impairments that occur with a silent delay were sufficient in our case to illuminate the underlying representation. Masking effects might nonetheless yield additional insights.

Relevance of f0-Based Pitch to Music. f0-Based pitch seems to be particularly important in music perception, evident in prior results documenting the effect of inharmonicity on music-related tasks. Melody recognition, “sour” note detection, and pitch interval discrimination are all worse for inharmonic than for harmonic tones, in contrast to other tasks such as up/down discrimination, which can be performed equally well for the two types of tones (shown again in the experiments here) (31). Our results here provide a potential explanation for these effects. Music often requires notes to be compared across delays or intervening sounds, as when assessing a note’s relation to a tonal center (56), and the present results suggest that this should necessitate f0-based pitch. Musical pitch perception may have come to rely on representations of f0 as a result, such that even in musical tasks that do not involve storage across a delay, such as musical interval discrimination, listeners use f0-based pitch rather than the spectrum (31). It is possible that similar memory advantages occur for other patterns that occur frequently in music, such as common chords.

Given its evident role in music perception, it is natural to wonder whether f0-based pitch is honed by musical training. Western-trained musicians are known to have better pitch discrimination than Western nonmusicians (57–59). However, previous studies examining effects of musicianship on pitch discrimination used either pure tone or harmonic complex tone stimuli, and thus do not differentiate between representations of the f0 vs. the spectrum. We found consistent overall pitch discrimination advantages for musicians compared to nonmusicians (*SI Appendix, Figs. S1–S3*), but found no evidence that this benefit was specific to f0 representations: Musicianship did not interact with the effects of inharmonicity or interstimulus delay. It is possible that more extreme variation in musical experience might show f0-specific effects. For instance, indigenous cultures in the Amazon appear to differ from Westerners in basic aspects of pitch (60) and harmony perception (61, 62), raising the possibility that they might also differ in the extent of reliance on f0-based pitch. It is also possible that musicianship effects might be more evident if memory were additionally taxed with intervening distractor tones.

Efficient Coding in Perception and Memory. Perception is often posited to estimate the distal causes in the world that generated a stimulus (63). Parameters that capture how a stimulus was generated are useful for behavior—as when one requires knowledge of an object’s shape in order to grasp it—but can also provide compressed representations of a stimulus. Indeed, efficient coding has been proposed as a way to estimate generative parameters of sensory signals (64). A sound’s f0 is one such generative parameter, and our results suggest that its representation may be understood in terms of efficient coding. Prior work has explained aspects of auditory representations (65, 66) and discrimination (67) as consequences of efficient coding, but has not explored links to memory. Our results raise the possibility that efficient coding may be particularly evident in sensory memory representations. We provide an example of abstract and compressed auditory memory representations, and in doing so explain some otherwise-puzzling results in pitch perception (chiefly, the fact that conventional pitch discrimination tasks are not impaired by inharmonicity).

This efficient coding perspective suggests that harmonic sounds may be more easily remembered because they are prevalent in the environment, such that humans have acquired representational transforms to efficiently represent them (e.g., by projection onto harmonic templates) (68). This interpretation also raises the possibility that the effects described here might generalize to or interact with other sound properties. There are many other regularities of natural sounds that influence perceptual grouping (69, 70). Each of these could in principle produce memory benefits when sounds must be stored across delays. In addition to regularities like harmonicity that are common to a wide range of natural sounds, humans also use learned “schemas” for particular sources when segregating streams of sound (38, 69) and these might also produce memory benefits. It is thus possible that recurring inharmonic spectral patterns, for instance in inharmonic musical instruments (32), could confer a memory advantage to an individual with sufficient exposure to them, despite lacking the mathematical regularity of the harmonic series.

Memory could be particularly important in audition given that sound unfolds over time, with the structures that matter in speech, music, and other domains often extending over many seconds. Other examples of auditory representations that discard details in favor of more abstract structures include the “contour” of melodies, which listeners retain in some conditions in favor of the exact f0 intervals between notes (71), or summary statistics of sound textures that average across temporal details (72, 73). These representations may reflect memory constraints involved in comparing two extended stimuli even without a pronounced interstimulus delay.

Future Directions. Our results leave open how the two representations implicated in pitch judgments are instantiated in the brain. Pitch-related brain responses measured in humans have generally not distinguished representations of the f0 from that of the spectrum (74–78), in part because of the coarse nature of human neuroscience methods. Moreover, we know little about how pitch representations are stored over time in order to mediate discrimination across a delay. In nonhuman primates, there is evidence for representations of the spectrum of harmonic complex tones (79) as well as of their f0 (80), although there is increasing evidence for heterogeneity in pitch across species (81–85). Neurophysiological and behavioral experiments with delayed discrimination tasks in nonhuman animals could shed light on these issues.

Our results also indicate that we unconsciously switch between representations depending on the conditions in which we must make pitch judgments (i.e., whether there is a delay between sounds). One possibility is that sensory systems can assess the reliability of their representations and base decisions on the representation that is most reliable for a given context. Evidence weighting according to reliability is a common strategy in perceptual decisions (86), and our results raise the possibility that such frameworks could be applied to understand memory-driven perceptual decisions.

Materials and Methods

Methods are described in full detail in *SI Appendix, SI Materials and Methods*. The full methods section includes descriptions of experimental participants and procedures, stimulus generation, data analysis, statistical tests, and power analyses.

Data Availability. All study data are included in the article and supporting information.

ACKNOWLEDGMENTS. We thank R. Grace, S. Dolan, and C. Wang for assistance with data collection, T. Brady for helpful discussions, and L. Demany, B. C. J. Moore, the entire J.H.M. laboratory, and two anonymous reviewers for comments on the manuscript. The work was supported by the McDonnell Foundation Scholar Award and NIH Grant R01DC014739 to J.H.M. and NIH Grant F31DC018433 and an NSF Graduate Research Fellowship to M.J.M.

1. F. C. Bartlett, *Remembering: A Study in Experimental and Social Psychology* (Cambridge University Press, 1932).
2. W. A. Phillips, On the distinction between sensory storage and short-term visual memory. *Percept. Psychophys.* **16**, 283–290 (1974).
3. Y. Jiang, I. R. Olson, M. M. Chun, Organization of visual short-term memory. *J. Exp. Psychol. Learn. Mem. Cogn.* **26**, 683–702 (2000).
4. T. F. Brady, T. Konkle, G. A. Alvarez, Compression in visual working memory: Using statistical regularities to form more efficient memory representations. *J. Exp. Psychol. Gen.* **138**, 487–502 (2009).
5. G. Sperling, The information available in brief visual presentations. *Psychol. Monogr.* **74**, 1–29 (1960).
6. M. I. Posner, S. W. Keele, Retention of abstract ideas. *J. Exp. Psychol. Gen.* **83**, 304–308 (1970).
7. M. I. Posner, “Abstraction and the process of recognition” in *Psychology of Learning and Motivation: Advances in Research and Theory*, J. T. Spence, G. H. Bower, Eds. (Academic, New York, 1969), vol. 3.
8. M. F. Panichello, B. DePasquale, J. W. Pillow, T. J. Buschman, Error-correcting dynamics in visual working memory. *Nat. Commun.* **10**, 3366 (2019).
9. C. J. Darwin, M. T. Turvey, R. G. Crowder, An auditory analogue of the Sperling partial report procedure: Evidence for brief auditory storage. *Cognit. Psychol.* **3**, 255–267 (1972).
10. N. Cowan, On short and long auditory stores. *Psychol. Bull.* **96**, 341–370 (1984).
11. W. A. Wickelgren, Distinctive features and errors in short-term memory for English consonants. *J. Acoust. Soc. Am.* **39**, 388–398 (1966).
12. W. A. Wickelgren, Distinctive features and errors in short-term memory for English vowels. *J. Acoust. Soc. Am.* **38**, 583–588 (1965).
13. D. B. Pisoni, Auditory short-term memory and vowel perception. *Mem. Cognit.* **3**, 7–18 (1975).
14. D. B. Pisoni, Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Percept. Psychophys.* **13**, 253–260 (1973).
15. B. H. Repp, Categorical perception: Issues, methods, findings. *Speech and Language* **10**, 243–335 (1984).
16. W. A. Wickelgren, Associative strength theory of recognition memory for pitch. *J. Math. Psychol.* **6**, 13–61 (1969).
17. D. W. Massaro, Retroactive interference in short-term recognition memory for pitch. *J. Exp. Psychol.* **83**, 32–39 (1970).
18. L. Demany, W. Trost, M. Serman, C. Semal, Auditory change detection: Simple sounds are not memorized better than complex sounds. *Psychol. Sci.* **19**, 85–91 (2008).
19. S. Clément, L. Demany, C. Semal, Memory for pitch versus memory for loudness. *J. Acoust. Soc. Am.* **106**, 2805–2811 (1999).
20. D. Deutsch, Mapping of interactions in the pitch memory store. *Science* **175**, 1020–1022 (1972).
21. C. Semal, L. Demany, Dissociation of pitch from timbre in auditory short-term memory. *J. Acoust. Soc. Am.* **89**, 2404–2410 (1991).
22. C. Semal, L. Demany, K. Ueda, P.-A. Hallé, Speech versus nonspeech in pitch memory. *J. Acoust. Soc. Am.* **100**, 1132–1140 (1996).
23. B. Tillmann, Y. Lévêque, L. Fornoni, P. Albouy, A. Caclin, Impaired short-term memory for pitch in congenital amusia. *Brain Res.* **1640**, 251–263 (2016).
24. T. Pasternak, M. W. Greenlee, Working memory in primate sensory systems. *Nat. Rev. Neurosci.* **6**, 97–107 (2005).
25. C. J. Plack, A. J. Oxenham, “The psychophysics of pitch” in *Pitch—Neural Coding and Perception*, C. J. Plack, A. J. Oxenham, R. R. Fay, A. J. Popper, Eds. (Springer Handbook of Auditory Research, Springer, New York, 2005), vol. 24, pp. 7–55.
26. A. de Cheveigne, “Pitch perception” in *The Oxford Handbook of Auditory Science*, C. J. P. Hearing, Ed. (Oxford University Press, New York, 2010), vol. 3.
27. American National Standards Institute, *American National Standard Acoustical Terminology, ANSI S1.1-1994* (American National Standards Institute, 1994).
28. B. C. J. Moore, B. R. Glasberg, Frequency discrimination of complex tones with overlapping and non-overlapping harmonics. *J. Acoust. Soc. Am.* **87**, 2163–2177 (1990).
29. A. Faulkner, Pitch discrimination of harmonic complex signals: Residue pitch or multiple component discriminations? *J. Acoust. Soc. Am.* **78**, 1993–2004 (1985).
30. C. Micheyl, K. Divis, D. M. Wroblewski, A. J. Oxenham, Does fundamental-frequency discrimination measure virtual pitch discrimination? *J. Acoust. Soc. Am.* **128**, 1930–1942 (2010).
31. M. J. McPherson, J. H. McDermott, Diversity in pitch perception revealed by task dependence. *Nat. Hum. Behav.* **2**, 52–66 (2018).
32. N. M. McLachlan, D. J. T. Marco, S. J. Wilson, The musical environment and auditory plasticity: Hearing the pitch of percussion. *Front. Psychol.* **4**, 768 (2013).
33. L. Demany, C. Ramos, On the binding of successive sounds: Perceiving shifts in non-perceived pitches. *J. Acoust. Soc. Am.* **117**, 833–841 (2005).
34. R. G. Crowder, Decay of auditory memory in vowel discrimination. *J. Exp. Psychol. Learn. Mem. Cogn.* **8**, 153–162 (1982).
35. L. Demany, C. Semal, J.-R. Cazalets, D. Pressnitzer, Fundamental differences in change detection between vision and audition. *Exp. Brain Res.* **203**, 261–270 (2010).
36. E. M. O. Borchert, C. Micheyl, A. J. Oxenham, Perceptual grouping affects pitch judgments across time and frequency. *J. Exp. Psychol. Hum. Percept. Perform.* **37**, 257–269 (2011).
37. T. R. Agus, S. J. Thorpe, D. Pressnitzer, Rapid formation of robust auditory memories: Insights from noise. *Neuron* **66**, 610–618 (2010).
38. K. J. P. Woods, J. H. McDermott, Schema learning for the cocktail party problem. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E3313–E3322 (2018).
39. T. M. Shackleton, R. P. Carlyon, The role of resolved and unresolved harmonics in pitch perception and frequency modulation discrimination. *J. Acoust. Soc. Am.* **95**, 3529–3540 (1994).
40. J. G. W. Bernstein, A. J. Oxenham, An autocorrelation model with place dependence to account for the effect of harmonic number on fundamental frequency discrimination. *J. Acoust. Soc. Am.* **117**, 3816–3831 (2005).
41. N. Cowan, J. S. Saults, L. D. Nugent, The role of absolute and relative amounts of time in forgetting within immediate memory: The case of tone-pitch comparisons. *Psychon. Bull. Rev.* **4**, 393–397 (1997).
42. C. Kaernbach, K. Schlemmer, The decay of pitch memory during rehearsal. *J. Acoust. Soc. Am.* **123**, 1846–1849 (2008).
43. R. McWalter, J. H. McDermott, Illusory sound texture reveals multi-second statistical completion in auditory scene analysis. *Nat. Commun.* **10**, 5096 (2019).
44. K. J. P. Woods, M. H. Siegel, J. Traer, J. H. McDermott, Headphone screening to facilitate web-based auditory experiments. *Atten. Percept. Psychophys.* **79**, 2064–2072 (2017).
45. J. B. Wilmer, How to use individual differences to isolate functional organization, biology, and utility of visual functions; with illustrative proposals for stereopsis. *Spat. Vis.* **21**, 561–579 (2008).
46. J. H. McDermott, A. J. Lehr, A. J. Oxenham, Individual differences reveal the basis of consonance. *Curr. Biol.* **20**, 1035–1041 (2010).
47. C. Spearman, The proof and measurement of association between two things. *Am. J. Psychol.* **15**, 72–101 (1904).
48. A. Bachem, Time factors in relative and absolute pitch discrimination. *J. Acoust. Soc. Am.* **26**, 751–753 (1954).
49. J. D. Harris, The decline of pitch discrimination with time. *J. Exp. Psychol.* **43**, 96–99 (1952).
50. S. J. Luck, E. K. Vogel, The capacity of visual working memory for features and conjunctions. *Nature* **390**, 279–281 (1997).
51. Y. Xu, Understanding the object benefit in visual short-term memory: The roles of feature proximity and connectedness. *Percept. Psychophys.* **68**, 815–828 (2006).
52. T. F. Brady, J. B. Tenenbaum, A probabilistic model of visual working memory: Incorporating higher order regularities into working memory capacity estimates. *Psychol. Rev.* **120**, 85–109 (2013).
53. S. Magnussen, M. W. Greenlee, R. Asplund, S. Dyres, Stimulus-specific mechanisms of visual short-term memory. *Vision Res.* **31**, 1213–1219 (1991).
54. V. A. Nemes, N. R. A. Parry, D. Whitaker, D. J. McKeefry, The retention and disruption of color information in human short-term visual memory. *J. Vis.* **12**, 26 (2012).
55. D. J. McKeefry, M. P. Burton, C. Vakrou, Speed selectivity in visual short term memory for motion. *Vision Res.* **47**, 2418–2425 (2007).
56. C. L. Krumhansl, *Cognitive Foundations of Musical Pitch* (Oxford Psychology Series, Oxford University Press, New York, 1990), vol. 17.
57. L. Kishon-Rabin, O. Amir, Y. Vexler, Y. Zaltz, Pitch discrimination: Are professional musicians better than non-musicians? *J. Basic Clin. Physiol. Pharmacol.* **12** (suppl. 2), 125–143 (2001).
58. C. Micheyl, K. Delhommeau, X. Perrot, A. J. Oxenham, Influence of musical and psychoacoustical training on pitch discrimination. *Hear. Res.* **219**, 36–47 (2006).
59. F. Bianchi, S. Santurette, D. Wendt, T. Dau, Pitch discrimination in musicians and non-musicians: Effects of harmonic resolvability and processing effort. *J. Assoc. Res. Otolaryngol.* **17**, 69–79 (2016).
60. N. Jacoby et al., Universal and non-universal features of musical pitch perception revealed by singing. *Curr. Biol.* **29**, 3229–3243.e12 (2019).
61. J. H. McDermott, A. F. Schultz, E. A. Undurraga, R. A. Godoy, Indifference to dissonance in native Amazonians reveals cultural variation in music perception. *Nature* **535**, 547–550 (2016).
62. M. J. McPherson et al., Perceptual fusion of musical notes by native Amazonians suggests universal representations of musical intervals. *Nat. Commun.* **11**, 2786 (2020).
63. D. Kersten, P. Mamassian, A. Yuille, Object perception as Bayesian inference. *Annu. Rev. Psychol.* **55**, 271–304 (2004).
64. F. Attneave, Some informational aspects of visual perception. *Psychol. Rev.* **61**, 183–193 (1954).
65. M. S. Lewicki, Efficient coding of natural sounds. *Nat. Neurosci.* **5**, 356–363 (2002).
66. N. L. Carlson, V. L. Ming, M. R. Deweese, Sparse codes for speech predict spectrotemporal receptive fields in the inferior colliculus. *PLoS Comput. Biol.* **8**, e1002594 (2012).
67. C. E. Stilp, T. T. Rogers, K. R. Kluender, Rapid efficient coding of correlated complex acoustic properties. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 21914–21919 (2010).
68. W. Młynarski, J. H. McDermott, Learning mid-level auditory codes from natural sound statistics. *Neural Comput.* **30**, 631–669 (2018).
69. A. S. Bregman, *Auditory Scene Analysis: The Perceptual Organization of Sound* (MIT Press, Cambridge, MA, 1990).
70. W. Młynarski, J. H. McDermott, Ecological origins of perceptual grouping principles in the auditory system. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 25355–25364 (2019).
71. W. J. Dowling, D. S. Fujitani, Contour, interval, and pitch recognition in memory for melodies. *J. Acoust. Soc. Am.* **49** (suppl. 2), 524 (1971).
72. J. H. McDermott, M. Schemitsch, E. P. Simoncelli, Summary statistics in auditory perception. *Nat. Neurosci.* **16**, 493–498 (2013).
73. R. McWalter, J. H. McDermott, Adaptive and selective time-averaging of auditory scenes. *Curr. Biol.* **28**, 1405–1418.e10 (2018).
74. R. D. Patterson, S. Uppenkamp, I. S. Johnsrude, T. D. Griffiths, The processing of temporal pitch and melody information in auditory cortex. *Neuron* **36**, 767–776 (2002).
75. H. Penagos, J. R. Melcher, A. J. Oxenham, A neural representation of pitch salience in nonprimary human auditory cortex revealed with functional magnetic resonance imaging. *J. Neurosci.* **24**, 6810–6815 (2004).

76. S. Norman-Haignere, N. Kanwisher, J. H. McDermott, Cortical pitch regions in humans respond primarily to resolved harmonics and are located in specific tonotopic regions of anterior auditory cortex. *J. Neurosci.* **33**, 19451–19469 (2013).
77. E. J. Allen, P. C. Burton, C. A. Olman, A. J. Oxenham, Representations of pitch and timbre variation in human auditory cortex. *J. Neurosci.* **37**, 1284–1293 (2017).
78. C. Tang, L. S. Hamilton, E. F. Chang, Intonational speech prosody encoding in the human auditory cortex. *Science* **357**, 797–801 (2017).
79. Y. I. Fishman, C. Micheyl, M. Steinschneider, Neural representation of harmonic complex tones in primary auditory cortex of the awake monkey. *J. Neurosci.* **33**, 10312–10323 (2013).
80. D. Bendor, X. Wang, The neuronal representation of pitch in primate auditory cortex. *Nature* **436**, 1161–1165 (2005).
81. J. K. Bizley, K. M. M. Walker, A. J. King, J. W. Schnupp, Neural ensemble codes for stimulus periodicity in auditory cortex. *J. Neurosci.* **30**, 5078–5091 (2010).
82. M. R. Bregman, A. D. Patel, T. Q. Gentner, Songbirds use spectral shape, not pitch, for sound pattern recognition. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 1666–1671 (2016).
83. X. Song, M. S. Osmanski, Y. Guo, X. Wang, Complex pitch perception mechanisms are shared by humans and a New World monkey. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 781–786 (2016).
84. K. M. M. Walker, R. Gonzalez, J. Z. Kang, J. H. McDermott, A. J. King, Across-species differences in pitch perception are consistent with differences in cochlear filtering. *eLife* **8**, e41626 (2019).
85. S. V. Norman-Haignere, N. Kanwisher, J. H. McDermott, B. R. Conway, Divergence in the functional organization of human and macaque auditory cortex revealed by fMRI responses to harmonic tones. *Nat. Neurosci.* **22**, 1057–1060 (2019).
86. M. O. Ernst, M. S. Banks, Humans integrate visual and haptic information in a statistically optimal fashion. *Nature* **415**, 429–433 (2002).