Harmonicity aids hearing in noise

Malinda J. McPherson[1,2], River C. Grace[1], and Josh H. McDermott[1,2,3,4]

1 Department of Brain and Cognitive Sciences, MIT
2 Program in Speech and Hearing Bioscience and Technology, Harvard University
3 McGovern Institute for Brain Research, MIT
4 Center for Brains, Minds and Machines, MIT

Abstract
Hearing in noise is a core problem in audition, and a challenge for hearing-impaired listeners, yet the underlying mechanisms are poorly understood. We explored whether harmonic frequency relations, a signature property of many communication sounds, aid hearing in noise. We measured detection thresholds in noise for tones and speech synthesized to have harmonic or inharmonic spectra. Harmonic signals were consistently easier to detect than otherwise identical inharmonic signals. Harmonicity also improved discrimination of sounds in noise. In contrast to other documented effects of harmonicity, harmonic detection advantages were comparable in musicians and non-musicians. The results show that harmonicity is critical for hearing in noise, demonstrating a previously unappreciated aspect of auditory scene analysis. The consistency of the effect across synthetic and natural stimuli, as well as across musical expertise, suggests its importance in everyday hearing.

Introduction

Noise is an unavoidable part of our auditory experience. We must pick out sounds of interest amid background noise on a daily basis – a speaker in a restaurant, a bird song in a windy forest, or a siren on a city street. Noise distorts the peripheral representation of sounds, but humans with normal hearing are relatively robust to its presence (Sarampalis et al., 2009). However, hearing in noise becomes more difficult with age (Ruggles et al., 2012; Tremblay et al., 2003) and for those with even moderate hearing loss (Bacon et al., 1998; Oxenham, 2008; Plack et al., 2014; Rossi-Katz & Arehart, 2005; Smoorenburg, 1992; Tremblay et al., 2003). Consequently, understanding the basis of hearing in noise, and its malfunction in hearing impairment, has become a major focus of auditory research (Kell & McDermott, 2019; Khalighinejad et al., 2019; Mesgarani et al., 2014; Moore et al., 2013; Rabinowitz et al., 2013; Town et al., 2019).

Hearing in noise can be viewed as a particular case of auditory scene analysis, the problem listeners solve when segregating individual sources from the mixture of sounds entering the ears (Bregman, 1990; Carlyon, 2004; Darwin, 1997; McDermott, 2009). In general, segregating sources from a mixture is possible only because of the regularities in natural sounds (Młynarski & McDermott, 2019). Most research on the signal properties that help listeners segregate sounds has focused on situations where people discern concurrent sources of the same type, for example, multiple speakers (the classic 'cocktail party problem' (Assman & Summerfield, 1990; Culling & Summerfield, 1995a; de Cheveigne, Kawahara, et al., 1997; de Cheveigne et al., 1995; de Cheveigne, McAdams, et al., 1997)), or multiple concurrent tones (as in music (Micheyl & Oxenham, 2010; Rasch, 1978)). Concurrent onsets or offsets (Darwin, 1981; Darwin & Ciocca, 1992), co-location in space (Cusack et al., 2004; Freyman et al., 2001; Hawley et al., 2004; Ihlefeld & Shinn-Cunningham, 2008), and frequency proximity (Chalikia & Bregman, 1993; Darwin & Hukin, 1997; Młynarski & McDermott, 2019) can all help to group sound elements and segregate them from other similar sounds in the background. Harmonicity – the property of frequencies that are multiples of a common 'fundamental', or f0 (Fig. 1) – likewise aids auditory grouping. For example, harmonic structure can help a listener select a single talker from a mixture of talkers (Darwin et al., 2003; Josupeit & Hohmann, 2017; Josupeit et al., 2020; Popham et al., 2018; Woods & McDermott, 2015). And when one harmonic in a complex tone or speech utterance is mistuned so that it is no longer an integer multiple of the fundamental, it can be heard as a separate sound (Hartmann et al., 1990; Moore et al., 1986; Popham et al., 2018; Roberts & Brunstrom, 1998).

Less is known about the factors and mechanisms that enable hearing in noise (operationally defined for the purposes of this paper as a background sound that does not contain audibly discrete frequency components, for example, white or pink Gaussian noise, and some sound textures). Previous research on hearing in noise has mainly focused on features of noise, such as stationarity, that aid its suppression (Kell & McDermott, 2019; Khalighinejad et al., 2019; Mesgarani et al., 2014; Moore et al., 2013; Rabinowitz et al., 2013) or separation (McWalter & McDermott, 2019; McWalter & McDermott, 2018) from signals such as speech. Here we instead study the aspects of a signal that enable it to be heard more readily in noise.

Harmonicity is one sound property that differentiates communication signals such as speech and music from noise. Although harmonicity is known to aid the segregation of multiple harmonically structured sounds, its role in hearing in noise is unclear. To explore whether harmonic frequency relations aid hearing of sounds in noise, we compared detection and discrimination of harmonic and inharmonic tones and spoken vowels embedded in noise. Inharmonic sounds were generated

by jittering frequency components so that they were not integer multiples of the fundamental frequency (McPherson & McDermott, 2018; Roberts & Holmes, 2006).

The first question was whether harmonicity would make sounds easier to detect in noise across a range of sounds and tasks. The one related prior study we know of found that 'chords' composed of three harmonically related pure tones were somewhat easier to detect in noise than non-harmonically related tones, but did not pursue the basis of this effect (Hafter & Saberi, 2001).

The second question was whether harmonicity would make sounds easier to discriminate in noise. To address this question, we measured pitch discrimination thresholds with harmonic and inharmonic stimuli at a range of signal-to-noise ratios (SNRs), asking whether harmonicity would aid discrimination in noise at supra-threshold SNRs. Pitch discrimination thresholds are known to be comparable for harmonic and inharmonic tones without noise, suggesting that listeners track the spectra of the tones in order to make up/down discrimination judgments (Faulkner, 1985; McPherson & McDermott, 2018, 2020; Micheyl et al., 2012; Moore & Glasberg, 1990). But in noisy conditions it could be difficult to accurately encode the entire spectrum, making it advantageous to rely on harmonic structure for discrimination. Previous studies have found that it is easier to hear the f0 of harmonic sounds when there is background noise (Hall & Peters, 1981; Houtgast, 1976), but it was unclear whether such effects would translate to improved discrimination of harmonic tones, compared to inharmonic tones, in noise.

A third question was whether any benefit of harmonicity would be influenced by musical experience. Musical training has been proposed as beneficial for hearing speech in noise (Clayton et al., 2016; Parbery-Clark et al., 2011; Swaminathan et al., 2015), but evidence for such musicianship advantages has been inconsistent (Boebinger et al., 2015; Madsen et al., 2019). It seemed plausible that musicianship effects might relate to harmonicity. Harmonic structure is critical to music – most musical instruments have harmonic frequency spectra, and the frequency ratios between common intervals in standard Western scales (and other scales around the world) are shared with the harmonic series. Musical training has been associated with the enhancement of perceptual judgments related to harmonicity, with lower pitch discrimination thresholds (Kishon-Rabin et al., 2001; McDermott, Keebler, et al., 2010; Micheyl et al., 2006; Spiegel & Watson, 1984), and larger preferences for harmonic over inharmonic sounds in musicians (Dellacherie et al., 2010; McDermott, Lehr, et al., 2010; Weiss et al., 2019), or in individuals with lifelong exposure to Western music (McDermott et al., 2016; McPherson et al., 2020). Thus, musical training might enhance sensitivity to harmonic structure, which, if relevant to hearing in noise, could support the use of musical training in clinical contexts.

We found that harmonic sounds were consistently easier to detect in noise than inharmonic sounds. This result held for speech as well as synthetic tones. Discrimination thresholds were also better for harmonic than inharmonic tones when presented in noise. This observed difference in discrimination thresholds between harmonic and inharmonic tones was greater than what could be explained simply by the detectability advantage for harmonic tones. Additionally, across both detection and discrimination tasks, we observed harmonic advantages for non-musicians on par with those for musicians. The results suggest a grouping effect in which harmonicity aids detection and discrimination of signals in noise.

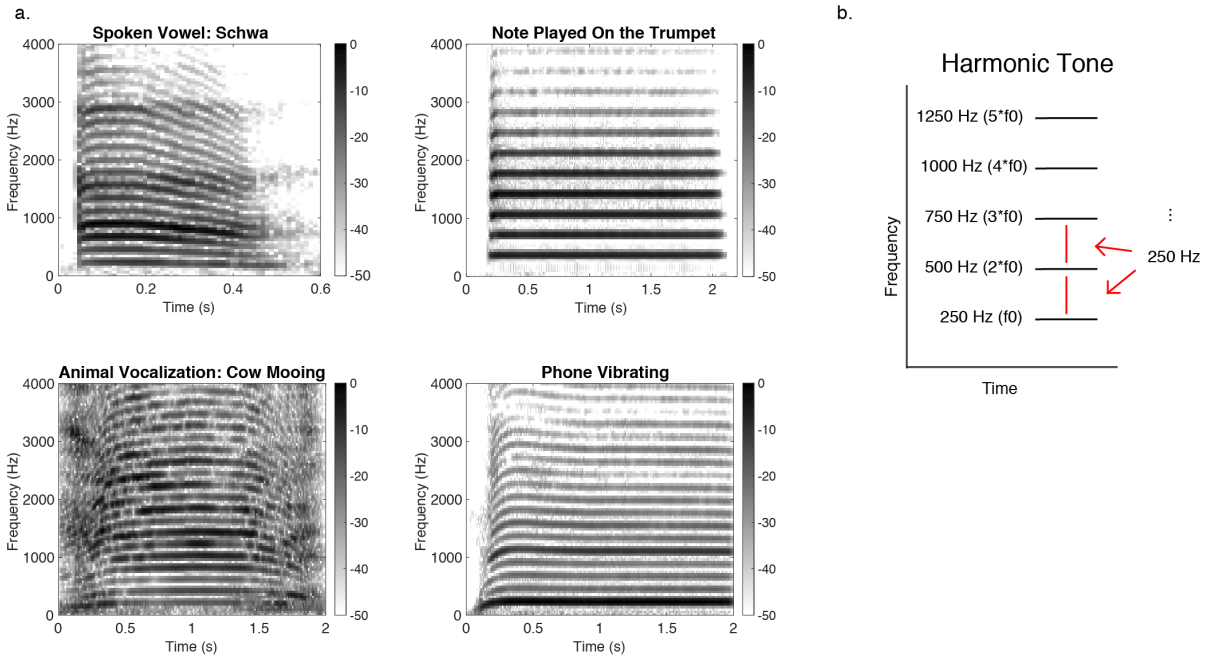


Figure 1. Harmonicity
a. Spectrograms of example natural harmonic sounds: a spoken vowel, a cow mooing, a note played on a trumpet, and a phone vibrating. The frequency components of such sounds are multiples of a fundamental frequency, and are thus regularly spaced across the spectrum. b. Schematic spectrogram of a harmonic tone with an f0 of 250 Hz.

Experiment 1. Detecting harmonic and inharmonic tones in noise.

The purpose of Experiment 1 was to examine the effect of harmonicity on the detection of sounds in noise. In each trial, participants heard two noise bursts (Fig. 1a). A complex tone (Fig. 1b.) or a pure tone was embedded in one of the noise bursts, and participants were asked to choose which noise burst contained the tone. The complex tones could be harmonic or inharmonic, with constituent frequencies added in sine or random phase. Participants completed four adaptive measurements of the detection threshold for tones in each condition.

Method

Participants. 100 participants completed Experiment 1 on an online data collection platform, Amazon Mechanical Turk. 14 of these participants were removed from analysis because their average threshold across conditions (using the first adaptive run of each condition) was over a standard deviation worse than the group mean across all conditions. This exclusion criterion is neutral with respect to the hypotheses being tested, and also independent of the data we analyzed (only the subsequent 3 runs were included for analysis in the remaining participants, to avoid double-dipping). Therefore, our exclusion procedure allowed unbiased threshold estimates from those final three runs. In previous studies we have found that online results replicate in-lab results when such steps are taken to exclude the worst-performing participants (McPherson & McDermott, 2020; Woods & McDermott, 2018). Of the remaining 86 participants (32 female, mean age = 38.6 years, S.D.=10.9 years), 39 had four or more years of musical training, with an average of 11.5 years, S.D.=8.9 years.

In this and other experiments, we determined sample sizes a priori based on pilot studies, and using G*Power (Faul et al., 2007). We ran a pilot study that was similar to Experiment 1. The

only difference was that the frequencies of each Inharmonic note were jittered independently on each trial (in contrast to Experiment 1, where each Inharmonic tone for a participant was made inharmonic in the same way across the entire study, as described below). We ran this pilot study in 43 participants, and observed a strong main effect of harmonicity (η_p^2=.37 for an ANOVA comparing Harmonic vs. Inharmonic conditions). We chose our sample size to be able to detect a potential musicianship effect that might be substantially weaker than the main effect of harmonicity. We thus sought to be well-powered to detect an interaction between musicianship and harmonicity 1/8 the size of the main effect of harmonicity at a significance level of $p<.01$, 95% of the time. This yielded a target sample size of 62 participants (31 musicians and 31 non-musicians). In practice we ran participants in batches, and then excluded them based on whether they passed the headphone check and our performance criteria, so the final sample was somewhat larger than this target.

Procedure. All experiments were conducted online on Amazon's Mechanical Turk platform. In-person data collection was not possible due to the COVID-19 virus. Prior to starting the experiment, potential participants were instructed to use headphones, and then used a calibration sound (1.5 seconds of Threshold Equalizing noise (Moore et al., 2000)) to set their volume to a comfortable level. The experimental stimuli were normalized to 6 dB below the level of the calibration sound to ensure that they were never uncomfortably loud. Participants were then screened with a brief experiment designed to help ensure they were wearing earphones or headphones, as instructed (Woods et al., 2017). If they passed this screening (across all experiments in this paper, approximately 65% of participants did so, consistent with Woods et. al, 2017), participants proceeded to the main experiment. For all experiments in the paper, participants received feedback after each trial, and to incentivize good performance, they received a compensation bonus proportional to the number of correct trials.

We used adaptive procedures to measure detection thresholds. Participants completed 3-down-1-up two-alternative-forced-choice ('does the first or second noise burst contain a tone?') adaptive threshold measurements. Adaptive tracks were stopped after 10 reversals. The signal-to-noise ratio (SNR) was changed by 8 dB for the first two reversals, 2 dB for the subsequent two reversals, and .5 dB for the final six reversals. The threshold estimate from a track was the average of the SNRs at the final six reversals. Participants completed four adaptive threshold measurements for each condition. Complex tone conditions (random vs. sine phase tones, and harmonic vs. inharmonic tones) were randomly intermixed, and the four runs of the pure tone condition were grouped together, run either before or after all of the complex tone adaptive runs, chosen equiprobably for each participant.

Stimuli. Trials consisted of two noise bursts, one of which contained a tone. First, two 900ms samples of noise were generated, and one of these noise samples was randomly chosen to contain the tone. The tone was scaled to have the appropriate power relative to that noise sample; both intervals were then normalized to the same rms value. Tones were 500ms in duration; the noise began and ended 200ms before and after the tone (Fig. 2a). The tones started 200ms after the noise to avoid an 'overshoot' effect, whereby tones are harder to detect when they start near the onset of noise. (Zwicker, 1965). The two noise bursts were separated by 200ms of silence.

The noise used in this and all other experiments was Threshold Equalizing (TE) noise (Moore et al., 2000). Pilot experiments with both white and pink noise suggested that the harmonic detection advantage is present regardless of the noise spectrum. In Experiment 1, noise was low pass filtered with a 6th order Butterworth filter to make it more pleasant for participants. The cutoff frequency was 6000 Hz, chosen to be well above the highest possible harmonic in the complex tones. Noise in all experiments was windowed in time with 10ms half-Hanning windows.

Complex tones contained ten equal-amplitude harmonics. Depending on the condition, harmonics were added in sine phase or random phase (Fig. 2c). F0s of the tones (both complex and pure – pure tones were just the f0 frequency component of the harmonic tones) were randomly selected to be between 200-267 Hz (log uniform distribution). Tones were windowed with 10ms half-Hanning windows, and were 500ms in duration. Tones and noise were sampled at 44.1 kHz.

To make tones inharmonic, the frequency of each frequency component (other than the f0 component) was 'jittered' by up to 50% of the f0 value. Jittering was accomplished by sampling a jitter value from the distribution U(-0.5, 0.5), multiplying by the f0, then adding the resulting value to the frequency of the respective harmonic. Jitter values were selected via rejection sampling, successively moving up the harmonic series rejecting or accepting sampled jitter values, to ensure that adjacent harmonics were always separated by at least 30 Hz (to avoid salient beating). Jitter values varied across participants (described below), but for a given participant were fixed across the experiment (i.e., each inharmonic tone heard by a given participant had the same jitter pattern).

For technical reasons all stimuli for online experiments were generated ahead of time. 20 stimuli were pre-generated for every possible difficulty level (SNR) within the adaptive procedure. The SNR was capped at +6 dB SNR per component. If participants in the experiment reached this cap the stimuli remained at this SNR until participants got three trials in a row correct. In practice, participants who performed poorly enough to reach this cap were removed post hoc by our filtering procedure. Adaptive tracks were initialized at -8 dB SNR per component. For each trial within an adaptive track, one of the 20 stimuli for the current difficulty level within the adaptive track was selected at random.

In order to vary the jitters across participants, we generated 20 independent sets of possible stimuli, each with a different set of randomly selected jitter values for the Inharmonic trials. Each participant only heard trials from one of these sets (i.e., all the inharmonic stimuli they heard were 'jittered' in the same way throughout the experiment). As some randomly selected jitter patterns can by chance be close to Harmonic, we randomly generated 100,000 possible jitter patterns, then selected the 20 patterns that minimized peaks in the autocorrelation function. The resulting 20 jitters were evaluated by eye to ensure that they were distinct.

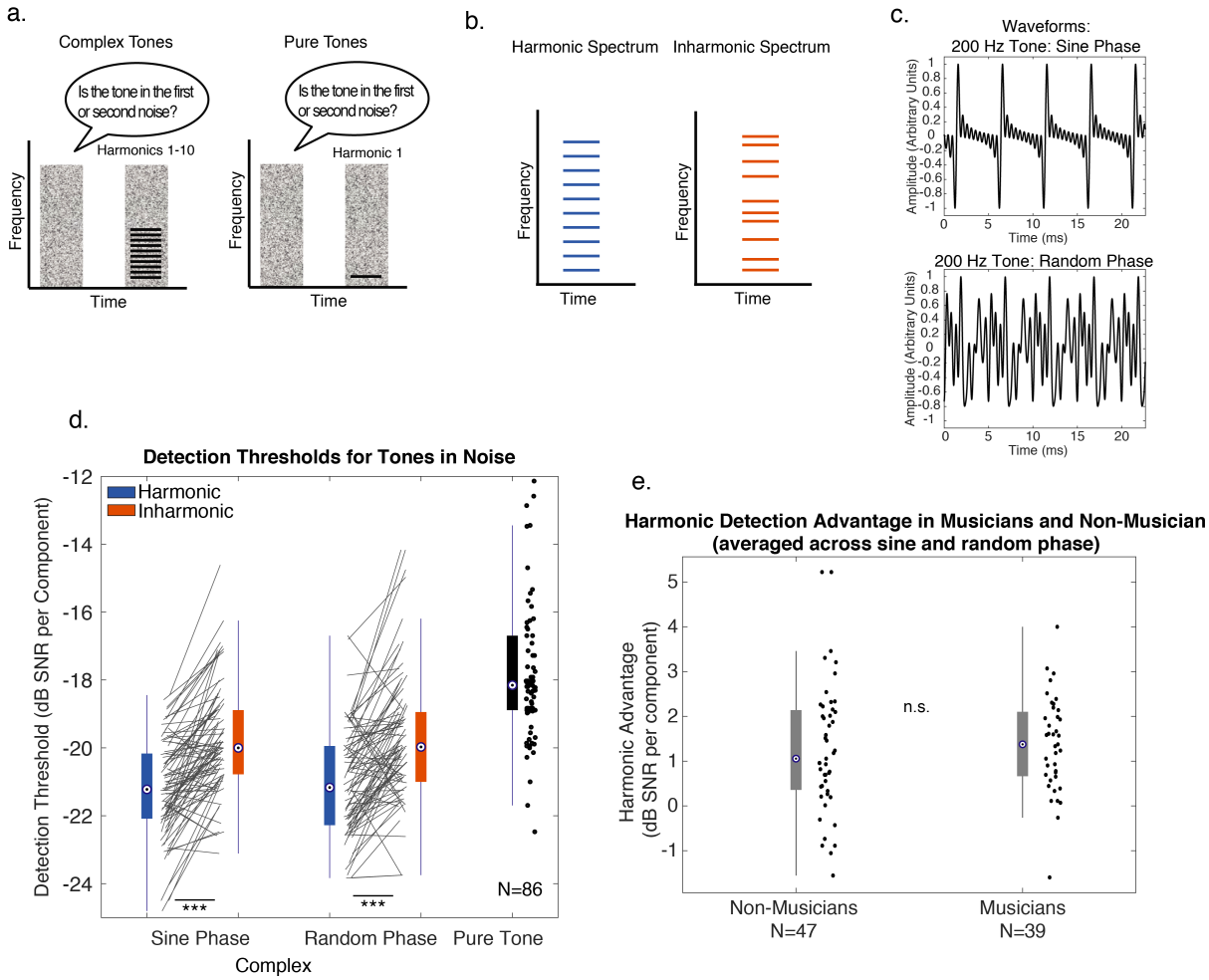


Figure 2: Harmonic advantage for detecting tones in noise
a. Trial structure for Experiment 1. During each trial, participants heard two noise bursts, one of which contained a complex tone (left) or pure tone (right), and were asked to decide whether the tone was in the first or second noise burst. b. Schematic of harmonic and inharmonic spectra; complex tones were harmonic or inharmonic (with frequencies jittered). c. Example waveforms of harmonic tones added in random phase (top) and sine phase (bottom). The waveform is 'peakier' when the harmonics are added in sine phase. d. Results of Experiment 1, shown as box-and-whisker-plots, with black lines for individual participant results. The central mark in the box plots the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. Whiskers extend to the most extreme data points not considered outliers. Asterisks denote significance, Wilcoxon signed-rank test: ***=p<0.001. e. Harmonic detection advantage (Harmonic threshold - Inharmonic threshold, averaged across phase conditions) for Musicians and Non-Musicians. For e, error bars denote standard error of the mean.

Statistical Analysis. Thresholds were calculated by averaging the SNR values of the final six reversals of the adaptive track. Data distributions were non-normal (skewed), so non-parametric tests were used for all comparisons. Wilcoxon signed-rank tests were used for pairwise comparisons between dependent samples. To compare performance across multiple conditions we used repeated-measures ANOVAs. However, because data were non-normal, we evaluated the significance of the F statistic with approximate permutation tests, randomizing the assignment of the data points across the conditions being tested 10,000 times, and comparing the F statistic to this distribution. To evaluate potential differences between musicians and non-musicians we compared the difference between harmonic and inharmonic detection thresholds. The distribution of these differences was normal (evaluated using the Lilliefors test), thus we used

independent sample t tests to compare groups, and used Bayesian statistics to estimate the probability of null results.

Results & Discussion

As shown in Fig. 2d, detection in noise was better for the complex tone conditions than the pure tone conditions (Z=6.49, p<.001, Pure Tone vs. mean performance for Inharmonic conditions) as expected from signal detection theory given the ten-fold increase in harmonics in the complex tones compared to the pure tones (Florentine et al., 1978). However, detection thresholds were substantially better for harmonic than inharmonic complex tones even though they each had 10 frequency components (significant differences in both sine and random phase conditions, Wilcoxon signed-rank test: sine phase: Z=6.97, p<.001; random phase: Z=6.83, p<.001). We observed a 2.38 dB SNR advantage for Inharmonic tones compared to Pure tones, and an additional 1.32 dB SNR advantage for Harmonic tones over Inharmonic tones (averaged across phase conditions).

These differences are large enough to have some real-life significance. For instance, if a harmonic tone could be just detected 10 meters away from its source in free field conditions, an otherwise identical inharmonic tone would only be audible 8.59 meters away from the source (using the inverse square law; for comparison, a pure tone at the same level as one of the frequency components from the complex tone would be audible 6.53 meters away).

A priori it seemed plausible that a detection advantage for harmonic tones could be explained by the regular amplitude modulation of harmonic sounds. However, performance was similar for the sine and random phase conditions (the latter of which produces substantially less modulation, Fig. 2b). We observed no significant differences between phase conditions or interaction with harmonicity (no significant main effect of phase, F(1,85)=1.07, p=.30,$\eta_p^{2=}$.012, and no interaction between harmonicity and phase, F(1,85)=0.89 p=.35, ,$\eta_p^{2=}$.01). This result indicates that the observed harmonic advantage derives from spectral rather than temporal properties of harmonic sounds.

The results are also unlikely to be explained by distortion products. Although harmonic tones would be expected to produce stronger distortion products than inharmonic tones, these should be undetectable for tones that include all the lower harmonics (as were used here and in most other experiments) (Norman-Haignere & McDermott, 2016; Pressnitzer & Patterson, 2001).

We were curious whether musicians and non-musicians would have comparable detection advantages for harmonic over inharmonic stimuli. Averaging across phase conditions, we compared the harmonic detection advantage for both groups ([Inharmonic thresholds - Harmonic thresholds]). We observed no significant differences between groups (Fig. 2e, and Supplementary Fig 1a, musician mean advantage = 1.28 dB, S.D.=1.46, non-musician mean advantage = 1.37 dB, S.D.-1.07, t(84)=0.32, p=.75). The Bayes factor (using a Cauchy distribution prior, centered at zero with a scale of .707) was .29 against a null hypothesis, providing moderate support for the null hypothesis (*JASP, Version 0.13.1*, 2020) .

Experiment 2. Discrimination thresholds in noise

In Experiment 2, we investigated whether the observed harmonic advantage would facilitate other types of judgments about sounds. Specifically, it seemed plausible that being better able to separate tones from noise might help listeners discriminate successive tones at low SNRs. Using an adaptive procedure, we measured traditional up-down discrimination thresholds for Harmonic, Inharmonic, and Pure Tone conditions (Fig. 3a) at a range of SNRs.

Method

Participants. 71 participants were recruited online for Experiment 2. We excluded participants who performed worse than 14.35% across all conditions (averaged across both runs of the experiment). This cutoff was based on a pilot study run in the lab – it was the average performance across all conditions for non-musician participants. We used this cutoff to obtain mean performance levels on par with those of compliant and attentive participants run in the lab. 27 participants were excluded from analysis by this criterion. This resulted in 44 participants, 15 female, mean age = 39.0 years, S.D.=10.75 years. 21 participants had four or more years of musical training, with an average of 11.1 years, S.D. = 12.1.

We chose our sample size using the same pilot data used to determine the exclusion criteria. The pilot experiment, run in 19 participants, differed from the current study in a few respects. In addition to being run in the lab, the pilot experiment did not include a pure tone condition, the SNR values were shifted half a semitone higher, and in Inharmonic conditions, a different jitter pattern was used for each trial (rather than the same jitter pattern being used across trials). We performed ANOVAs testing for effects of harmonicity and musicianship; the pilot data showed fairly large main effects of both harmonicity (η_p^2=.77) and musicianship (η_p^2=.45), suggesting that both these analyses would be well-powered with modest sample sizes. To ensure the reliability of planned analyses examining the inflection points and slopes of sigmoid functions fitted to the discrimination curves, we also estimated the sample size needed to obtain reliable mean thresholds. We extrapolated from our pilot data (via bootstrap) that an N of at least 36 would be necessary to have a split-half reliability of the mean measured threshold in each condition (assessed between the first and second adaptive runs of the experiment) greater than r=.95. This sample size was also sufficient for the ANOVA analyses (for example, to see an effect of musicianship 1/2 the size of that observed in our pilot study 95% of the time at a p<.01 significance level, one would need a sample size of 28). We thus aimed to recruit at least 36 participants.

Procedure. In Experiment 2 we measured classic up-down "pitch" discrimination. As in Experiment 1, on each trial participants heard two noise bursts. However, in this experiment, a tone was presented in each of the two noise bursts, and participants judged whether the second tone was higher or lower than the first tone. The difference in the f0s used to generate the tones was initialized at 1 semitone, and was changed by a factor of 2 through the first four reversals, and then by $\sqrt{2}$ through the final six reversals. We tested pitch discrimination at 6 SNRs for pure tones, 7 SNRs for inharmonic tones, and 8 SNRs for harmonic tones. This choice was motivated by pilot data showing that at the lowest SNR tested for harmonic tones, inharmonic tones were undetectable. The same logic applied to the lowest two SNRs and pure tones. Because we expected that the lowest SNR conditions tested in each condition would make discrimination very difficult (if not impossible), we capped the f0 difference at 4 semitones. If participants completed three trials in a row incorrectly at this f0 difference, the adaptive track was ended early. For these trials, the threshold was conservatively recorded as 4 semitones for analysis (25.99%). Participants performed 2 adaptive runs per condition.

Stimuli: The stimuli for Experiment 2 were identical to the random phase complex tones used in Experiment 1, except that each of the two noise bursts contained a tone. The initial f0 for each trial was randomly selected between 200 and 267 Hz (log uniform distribution). The same vector of jitter values was applied to each of the two notes (McPherson & McDermott, 2018) used in a trial. As in Experiment 1, we generated 20 sets of stimuli, each with a different jitter pattern, selected from 100,000 randomly generated jitter patterns as those with the smallest autocorrelation peaks. Each participant was randomly assigned one of these sets of stimuli, and only heard one inharmonic 'jitter' pattern throughout the experiment.

Statistical Analysis. Thresholds were estimated by taking the geometric mean of the f0 differences from the final six reversals of the adaptive track. As in Experiment 1, data distributions were non-normal (skewed), so Wilcoxon signed-rank tests were used for pairwise comparisons. To compare performance across multiple conditions or across musicianship we used repeated-measures ANOVAs (for within group effects) and mixed-model ANOVAs (to compare within and between group effects). We evaluated the significance of the F statistic with approximate permutation tests, randomizing the assignment of the data points across the conditions being tested 10,000 times, and comparing the F statistic to this null distribution.

We completed a secondary analysis to compare the results for the three stimulus conditions (Harmonic, Inharmonic, and Pure Tone) after accounting for differences in detectability between conditions. We replotted the pitch discrimination curves in terms of the SNR relative to the detection thresholds measured in Experiment 1 (-21.00 dB SNR, -19.77 dB SNR, -17.39 dB SNR, for Harmonic, Inharmonic and Pure Tone conditions, respectively, Fig. 2b, inset). We then bootstrapped over participants to evaluate the statistical significance of the residual differences between conditions. For each bootstrap sample we fit a sigmoid function to the results curve for each condition and compiled a distribution of the slopes and midpoints of each of the curves. In order to obtain reasonable curve fits, it was necessary to pad the data on either end of the SNR range with dummy values: with -25.99 on the low end (the highest possible threshold that could be measured in the experiment, as if we had added one additional, lower SNR), and with zeros at the high end. We compared the bootstrap distributions of the slopes and midpoints of the three conditions to determine the significance of differences between conditions.

Results & Discussion

Replicating prior results (McPherson & McDermott, 2018, 2020), discrimination thresholds in quiet were similar for harmonic and inharmonic tones (around 1.5% in both cases; rightmost conditions of Fig. 3b), with indistinguishable thresholds (-Inf dB; $Z=1.19$, $p=.23$). However, at lower SNRs inharmonic discrimination thresholds were substantially higher than harmonic thresholds (significant differences at all SNRs between -20.5 and -13 dB; $Z>2.61$, $p<.009$ in all cases), yielding a main effect of harmonicity (between Harmonic and Inharmonic tones (excluding -22 dB SNR, when only Harmonic thresholds were measured), $F(6,43)=171.9$, $p<.0001$, $\eta_p^2=.80$).

When we accounted for differences in the detection thresholds for the three tone types (as measured in Experiment 1), we found that the inflection points of the sigmoid functions remained significantly different for Harmonic and Inharmonic conditions ($p=.036$). The inflection point for the Pure Tone condition was not significantly different from either the Harmonic ($p=.99$) or Inharmonic conditions ($p=.68$), and the slopes of the three conditions were not significantly different. The difference between Harmonic and Inharmonic discrimination after accounting for the detectability of the tones suggests that harmonic discrimination advantage is better than what would be expected based on detectability, or conversely, that inharmonic discrimination is worse than what would be expected based on detectability. Even at SNRs where people can detect inharmonic tones fairly reliably, harmonicity aids discrimination in noise, perhaps because representations of the f0 can be used for discrimination.

Consistent with many previous studies (Kishon-Rabin et al., 2001; McDermott, Keebler, et al., 2010; Micheyl et al., 2006; Spiegel & Watson, 1984), pitch discrimination was overall better in musicians than non-musicians (Supplementary Fig. 1b). These differences were significant by a sign test (mean thresholds were higher in non-musicians for 19 of 21 conditions, $p<.001$), but were modest, and did not reach significance in an ANOVA (excluding the -22 and -20.5 dB SNR conditions, for which we didn't measure Pure Tone thresholds, $F(1,42)=1.66$, $p=.20$, $\eta_p^2=.04$). We also did not observe a significant interaction between musicianship and Harmonicity (only examining the Harmonic and Inharmonic conditions, -20.5 dB SNR and greater, $F(1,42)=0.04$,

p=.84, η_p^2=.001). It is possible that if we increased the level of musical training used to categorize participants as musicians, we would have seen a more robust difference in discrimination ability between groups. However, our results show that the Harmonic advantage for pitch discrimination is observed in both musicians and non-musicians (significant main effect in each group separately, musicians: F(6,20)=123.16, p<.0001, η_p^2=.86, non-musicians F(6,22)=70.55, p<.0001, η_p^2=.76), which suggests that this advantage reflects a domain-general auditory scene analysis ability.

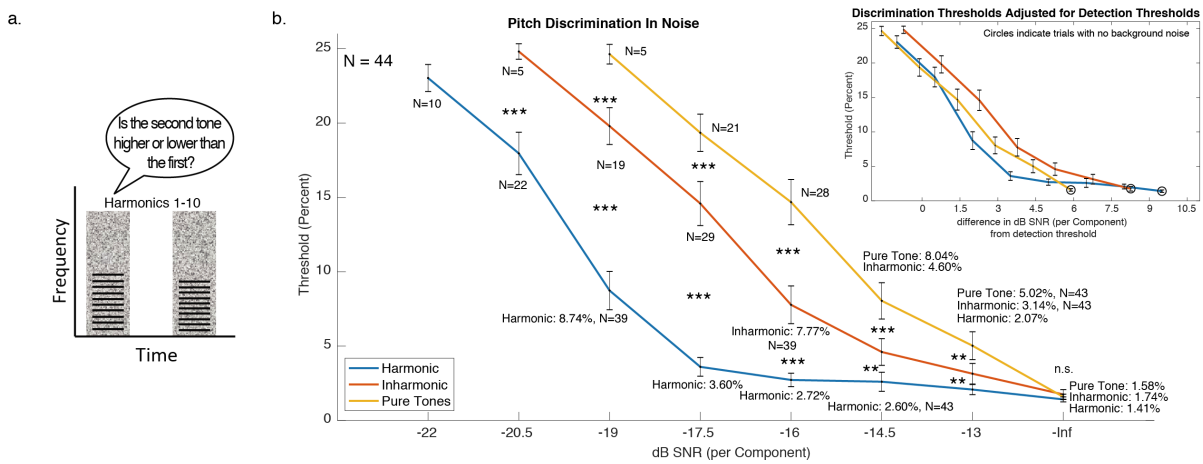


Figure 3: Harmonic advantage for discriminating tones in noise

a. Schematic of the trial structure for Experiment 2. During each trial, participants heard two noise bursts, each of which contained a complex tone (both tones were either harmonic or inharmonic), and were asked to decide whether the second tone was higher or lower than the first tone. b. Results from Experiment 2. Error bars denote standard error of the mean. For conditions where we were unable to measure thresholds from all participants, the number of participants with measurable thresholds is indicated next to the data point. Exact threshold values are provided for thresholds under 10%. Asterisks denote significance, Wilcoxon signed-rank test: ***=p<0.001, **=p<0.01. Inset: Pitch discrimination thresholds adjusted based on the detection thresholds measured in Experiment 1. The x axis plots SNR relative to detection threshold for the three different tone types.

Experiment 3. Detecting speech in noise

In Experiment 3, we tested whether the harmonic detection advantage would be present for natural sounds by measuring detection thresholds for spoken syllables embedded in noise (Fig. 4a).

Method

Participants: 78 participants completed Experiment 3 online. 9 were removed because their average performance across the first run of all conditions was over a standard deviation away from the group mean across the first run. As in Experiment 1, only the subsequent 3 runs were used for analysis. 69 participants were included in the final analysis, 33 female, mean age=36.9 years, S.D.=11.9 years). 20 participants had four or more years of musical training, with an average of 10.9 years, S.D.=9.3 years.

The effect size of harmonicity measured in a pilot version of Experiment 3 was moderate (d_z = 0.39), plausibly because we used natural stimuli, which are more variable than the synthetic tones used in other experiments (in which we observed larger effect sizes). The pilot experiment (run in 125 participants) was identical to Experiment 3 except that each Inharmonic trial contained harmonics that were jittered independently from the other trials. We aimed to run at least 61

participants to be 90% sure of seeing an effect this size with a .05 significance threshold. We did not attempt to recruit equal numbers of musicians and non-musicians given the lack of an effect of musicianship in Experiment 1.

Procedure. We measured detection thresholds for single spoken vowels embedded in noise, resynthesized to be inharmonic or harmonic. Participants were asked whether the first or second noise burst contained a word, rather than a tone. The adaptive staircase procedure was the same as that used in Experiment 1.

Stimuli: Speech was resynthesized using the STRAIGHT analysis and synthesis method (Ellis et al., 2012; Kawahara & Morise, 2011). STRAIGHT decomposes a recording of speech into voiced and unvoiced vocal excitation and vocal tract filtering. If the voiced excitation is modelled sinusoidally, one can alter the frequencies of individual harmonics and then recombine them with the unaltered unvoiced excitation and vocal tract filtering to generate inharmonic speech. This manipulation leaves the spectral envelope of the speech largely intact, and intelligibility of inharmonic speech in quiet is comparable to that of harmonic speech (Popham et al., 2018). The frequency jitters for inharmonic speech were chosen in the same way as those for inharmonic complex tones. Speech and noise were sampled at 24kHz. Code implementing the harmonic/inharmonic resynthesis is available on the senior author's lab web page. We used the vowels /i/, /u/, /a/ and /ɔ/, from the Hillenbrand vowel set (Hillenbrand et al., 1995) (h-V-d syllables). These four vowels bound the English vowel space.

As in Experiment 1 participants heard syllables embedded in TE-noise, but noise bursts were 650ms in duration and vowels were truncated to be 250ms in duration. There were 200ms of noise before the onset of the syllable and 200ms of noise after the syllable ended. Noise was not low-pass filtered; noise went up to the Nyquist limit.

Stimuli were pre-generated, and 20 trials were generated in advance for each possible difficulty level. The adaptive procedure was initialized at an SNR of 2 dB SNR and capped at 16 dB SNR. The same pattern of jitter was used throughout the entire speech syllable, and as in previous studies, 20 different sets of stimuli were generated, each with a distinct jitter pattern for inharmonic stimuli. Participants were randomly assigned to one of the 20 stimuli sets.

Statistical Analysis. As in Experiment 1, thresholds were calculated by averaging the SNR values of the final six reversals of the adaptive track. Data distributions were non-normal (skewed), so a Wilcoxon signed-rank test was used to compare the Harmonic and Inharmonic conditions.

Results & Discussion

As shown in Fig. 4b, harmonic vowels were easier to detect in noise than inharmonic vowels (Wilcoxon signed-rank test: $Z=3.61$, $p<.001$). This result demonstrates that the effect observed with complex tones generalizes to real-world sounds such as speech. Perhaps unsurprisingly, the harmonic advantage was more variable with these natural stimuli (the standard deviation of the difference between harmonic and inharmonic thresholds was 1.29 dB SNR in Experiment 1, but 2.32 dB SNR in Experiment 3). This variability may reflect the additional cues available for detection in some stimulus exemplars but not others, including concurrent modulation across frequency components (Culling & Summerfield, 1995b; McAdams, 1989), and onsets and offsets of consonants (Darwin, 1981). The persistence of the harmonic advantage despite these factors suggests that the advantage for detecting harmonic sounds could aid in real-world listening.

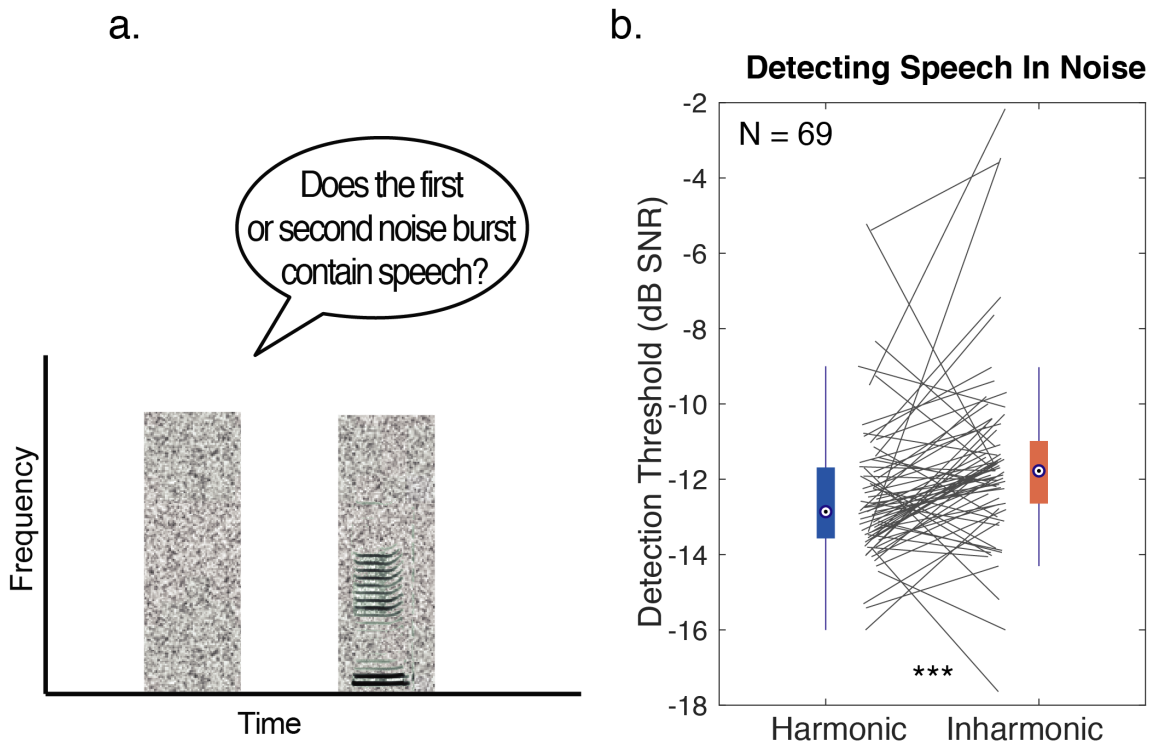


Figure 4: Harmonic advantage for detecting speech in noise
a. Schematic of the trial structure for Experiment 3. During each trial, online participants heard two noise bursts, one of which contained a spoken syllable, and were asked to decide whether the first or second noise burst contained speech. Speech was resynthesized to be either harmonic or inharmonic. b. Results of Experiment 3. Results are shown as box-and-whisker-plots, with black lines plotting individual participant results. The central mark in the box plots the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. Whiskers extend to the most extreme data points not considered outliers. Asterisks denote significance, Wilcoxon signed-rank test: ***=p<0.001.

Experiment 4. Detecting harmonic and inharmonic tones without resolved harmonics

Due to the increase in cochlear filtering bandwidth with frequency, only harmonics below about the 10th are believed to be individually discernible by the auditory system, and these harmonics dominate the perception of pitch (Shackleton & Carlyon, 1994). To determine whether the harmonic detection advantage observed in Experiment 1 was driven by low-numbered harmonics that are individually "resolved" by the cochlea, we ran a second experiment with the same task as Experiment 1, but with tones filtered to only contain harmonics 12-21 ("unresolved" harmonics, Fig. 5a). Tones were again presented in either sine phase or random phase.

Method

Participants. 30 participants were recruited online for Experiment 4. 6 participants performed over a standard deviation worse than the group mean on the first adaptive run and were excluded from analysis. As in previous studies, only the subsequent 3 runs were analyzed. 24 participants were included in the final analysis, 8 female, with a mean age of 36.4 years, S.D.=9.9 years. 10 participants had four or more years of musical training, with an average of 11.1 years, S.D.=9.7 years.

We used the pilot data from Experiment 1 to determine sample size. Based on prior work we hypothesized that the effect of harmonicity might be reduced with unresolved harmonics (Hartmann et al., 1990; Moore et al., 1985). We aimed to be able to detect an effect half the size of the main effect of harmonicity seen with resolved harmonics in our pilot data (η_p^2=.37). This yielded a target sample size of 15 participants (to have a 95% chance of seeing the hypothesized effect with a .01 significance threshold).

Procedure. The instructions and adaptive procedure were identical to those used in Experiment 1.

Stimuli. Tones contained harmonics 12 to 21 at full amplitude, with a trapezoid-shaped filter applied in the frequency domain in order to reduce the sharp spectral edge that might otherwise be used to perform the task. On the lower edge of the tone, the 10th harmonic was attenuated to be 30 dB below the 12th harmonic, and the 11th harmonic to be 15 dB below. On the upper edge of the tone, the same pattern of attenuation was applied in reverse between the 21st and 23rd harmonics. All other harmonics were removed. Additionally, the cutoff frequency for the noise (TE-noise) was increased to 10,000 Hz (rather than 6,000 Hz used in Experiments 1, 2, and 5). Noise was filtered with a 6th order Butterworth filter. Other aspects of the stimuli (duration of tones, timing of tones in noise, etc.) were matched to parameters used in Experiment 1.

Statistical Analysis. Statistical analysis was identical to that used in Experiment 1.

Results and Discussion

As shown in Fig. 5b, there was no difference in detectability between harmonic and inharmonic stimuli ($F(1,23)=1.93$, $p=.12$, $\eta_p^2=.08$). There was also no main effect of phase ($F(1,23)=0.99$, $p=0.33$, $\eta_p^2=.04$). These results suggest that the harmonic detection advantage is specific to resolved harmonics, placing constraints on the mechanism underlying the harmonic advantage observed in our other experiments. The Bayes factor (using a Cauchy distribution prior, centered at zero with a scale of .707) was .34 against a null hypothesis, providing mild support for the null hypothesis (*JASP, Version 0.13.1*, 2020).

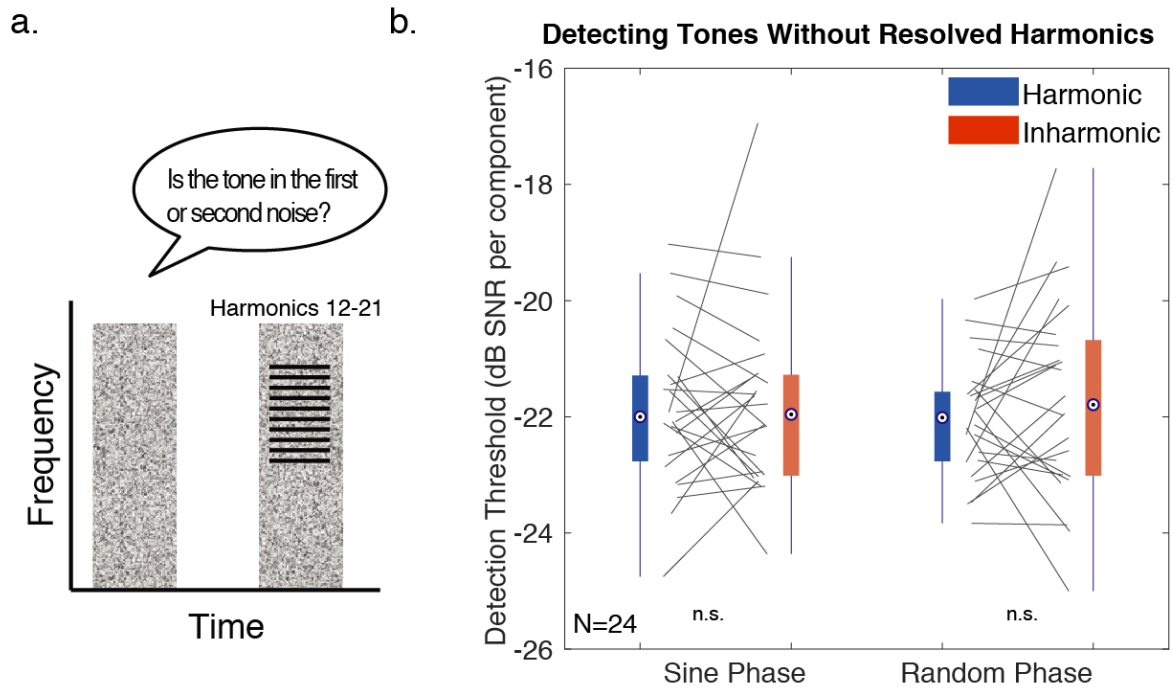


Figure 5: Harmonic detection advantage is specific to resolved harmonics
a. Schematic of the trial structure for Experiment 4. During each trial, participants heard two noise bursts, one of which contained a complex tone with unresolved harmonics, and were asked to decide whether the first or second noise burst contained a word. b. Results from Experiment 4, shown as box-and-whisker-plots, with black lines plotting individual participant results. The central mark in the box plots the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. Whiskers extend to the most extreme data points not considered outliers.

Experiment 5. Detecting harmonic and inharmonic tones in noise, with cueing

One potential explanation for the observed harmonic detection advantage is that people have internal templates for harmonic spectra, potentially developed over a lifetime of exposure to harmonic sounds to help efficiently encode natural sounds. These templates could potentially help listeners know what to listen for in a detection task. Experiment 5 tested this idea by assessing whether the harmonic advantage persists even when listeners are cued beforehand to the target tone. Participants heard two stimulus intervals, each containing a “cue” tone followed by a noise burst. One of the noise bursts contained an additional occurrence of the cue tone (Fig. 6a), and participants were asked whether the first or second noise burst contained the cued tone.

Method

Participants. 29 participants completed Experiment 5 online. 5 participants were removed because their average performance across the first run of both conditions was over a standard deviation lower than the group mean. As in previous experiments, only the subsequent 3 runs were used for analysis. 24 participants were included in the final analysis, 10 female, mean age=38.4, S.D.=11.8 years. 11 participants had four or more years of musical training, with an average of 10.2 years, S.D.=4.7 years.

We used data from a pilot experiment to determine sample size. The pilot experiment differed from the current experiment in 2 ways: it was run in the lab, and each Inharmonic note contained harmonics that were jittered independently from the other trials. The pilot experiment was run in 17 participants. Since Experiment 5 only had two conditions, we intended to use a single Wilcoxon signed-rank test to assess the difference between Harmonic and Inharmonic conditions. The effect size for this comparison in the pilot experiment was $d_z = .76$. This yielded a target sample size of at least 18 participants to have a 90% chance of seeing an effect of harmonicity of that size with a .05 significance threshold using a Wilcoxon signed-rank test.

Procedure. The instructions and adaptive procedure were identical to those used in Experiment 1.

Stimuli. Participants heard a tone before each of the two noise bursts. This "cue" tone was identical to the tone embedded in one of the noise bursts (that participants had to detect). Each trial had the following structure: a 500ms tone, followed by 200ms of silence, the first 900ms noise burst, 400ms of silence, a 500ms tone, 200ms of silence, and finally, the second 900ms noise burst. The target tone was present in either the first or the second noise burst, starting 200ms into the noise burst and lasting for 500ms. Only tones with harmonics added in random phase were used. In all other respects, stimuli in Experiment 5 were identical to those of Experiment 1.

Statistical Analysis. Thresholds were calculated by averaging the SNR values of the final six reversals of the adaptive track. Wilcoxon signed-rank tests was used to compare the Harmonic and Inharmonic conditions.

Results & Discussion

As shown in Fig. 6b, the harmonic advantage persisted with the cue ($Z=3.17$, $p=.002$). Even when participants knew exactly what to listen for in the noise, there was still an added benefit when detecting harmonic tones. Wilcoxon rank-sum tests showed that detection thresholds for Harmonic tones in Experiment 1 (without a cue) were indistinguishable from those with a cue tone ($Z=1.73$, $p=.08$). There was also no significant difference for Inharmonic tones with and without a cue ($Z=0.82$, $p=0.41$). This result suggests that the observed detection advantage for Harmonic over Inharmonic tones does not simply reflect biases due to familiarity.

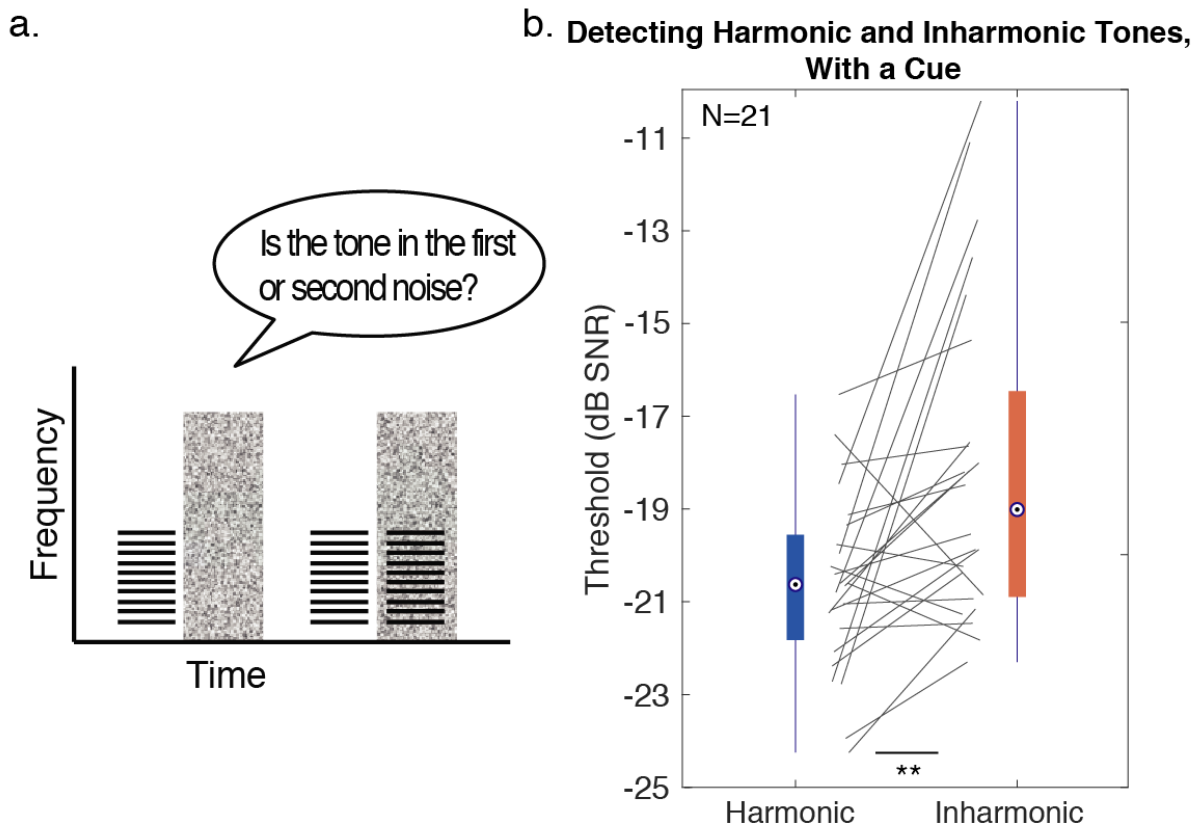


Figure 6: Harmonic advantage persists when listeners know what to listen for
a. Schematic of the trial structure for Experiment 5. During each trial, participants heard two noise bursts, both of which were preceded by a 'cue' tone, and one of which contained that same 'cue' tone. Participants were asked to decide whether the first or second noise burst contained the cued tone. b. Results from Experiment 5. shown as box-and-whisker-plots, with black lines plotting individual participant results. The central mark in the box plots the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. Whiskers extend to the most extreme data points not considered outliers. Asterisks denote significance, Wilcoxon signed-rank test: **=p<0.01.

General Discussion

We found that both synthetic and natural harmonic sounds were consistently easier to detect and discriminate in noise than otherwise identical inharmonic sounds. This harmonic advantage was present in both musicians and non-musicians, and generalized to natural sounds (specifically, speech, as seen in Experiment 3). Most acoustic communication signals (e.g. speech, and many musical sounds) are harmonic. The benefit we observed suggests that such harmonic sounds are audible at greater distances than otherwise identical inharmonic sounds. For example, if a harmonic sound is just detectable 10 meters from its source in a scene with spatially uniform background noise, our results indicate that the listener would have to move 1.4 meters closer to the sound source to hear an otherwise identical inharmonic sound. Given the ubiquity of both background noise and harmonic communication sounds in daily life, humans may have evolved or learned mechanisms to help detect harmonic sounds in noisy backgrounds. These results represent a previously undocumented aspect of auditory scene analysis.

Perceptual basis for harmonic detection advantage in noise

The results of Experiment 1 suggest that the harmonic advantage cannot be explained by amplitude fluctuations in the sound wave (because results were similar for sine and random phase conditions). Moreover, our results suggest that the harmonic advantage is absent for tones containing only unresolved harmonics, in which amplitude fluctuations should be maximally salient. The harmonic detection advantage persisted even when people knew exactly what to listen for (Experiment 5), suggesting that it is perceptual in origin, and reflects a basic ability to separate harmonic sounds from noise.

Advantages for pitch discrimination

The observation that harmonicity enhances pitch discrimination in noise (Experiment 2) suggests that harmonicity aids the ability to extract information from sounds in noise, in addition to improving detectability. Harmonic and inharmonic discrimination at high SNRs is similar, likely driven by the ability to track frequency shifts between notes (Demany & Ramos, 2005; Faulkner, 1985; McPherson & McDermott, 2018; Micheyl et al., 2012; Moore & Glasberg, 1990). But when presented in background noise, harmonic structure may help listeners hear out individual harmonics, or alternatively, enable listeners to 'fill in' missing harmonics that they would otherwise be unable to hear (McDermott & Oxenham, 2008). When notes are inharmonic, if a listener fails to detect even one of the frequency components from the complex, it could make the correspondence between the components of the first note and the second note (and thus the direction of the pitch change) ambiguous, impairing performance.

Could the harmonic discrimination advantage reflect f0-based pitch? Previous studies suggest that listeners are more inclined to base judgments on the f0, rather than spectral features, when tones are embedded in noise (Hall & Peters, 1981; Houtgast, 1976). Thus, in principle, the detection advantage we report could be driven by a pitch signal from harmonically related frequencies. Some evidence for this possibility comes from the analysis of Experiment 2 that adjusted discrimination thresholds using detection thresholds. The harmonicity advantage persisted even after adjusting for the detection advantage, suggesting that there is some additional factor (potentially f0-based pitch) that aids discrimination when sounds are harmonic. This result is consistent with the idea that the grouping of harmonics and the estimation of their f0 may rely on partially distinct mechanisms (Brunstrom & Roberts, 2000; Moore, 1987; Roberts & Brunstrom, 2001).

Musicianship effects

Prior experiments comparing sound segregation in musicians and non-musicians have reached divergent conclusions about the benefits of musical training (Madsen et al., 2019). Studies that show a musician benefit have often involved multiple harmonic sounds, or a single harmonic mistuned from a complex tone (Coffey et al., 2017; Zendel & Alain, 2009). In contrast, we found musical training to have little effect on the harmonic advantage for detecting sounds in noise. Harmonic detection advantages were comparable in musicians and non-musicians, as was the effect of noise on discrimination. Western musical training often involves hearing out one harmonic sound among others (picking out a melody from its harmony, for example), but does not typically require hearing in stationary background noise of the sort used here. It is thus possible that hearing a harmonic sound in noise requires distinct strategies compared to hearing out one harmonic sound among other similar sounds, and that these two abilities could be differentially affected by musicianship (Oxenham et al., 2003). Musicianship advantages might thus be expected in conditions where there are multiple harmonic sounds, but not in conditions like those tested here, where a single tone is embedded in noise.

Future directions

Our findings suggest that harmonicity is critical for detecting and discriminating sounds in noisy auditory scenes. In contrast to traditional experiments measuring speech intelligibility in noise (Duquesnoy, 1983; Festen & Plomp, 1990), we demonstrate a hearing-in-noise effect with relatively simple stimuli. Our effects might plausibly be evident in non-human animal models of hearing (Feng & Wang, 2017), and could be used to further explore and understand cross-species similarities and differences in the representations of harmonic sounds (Kalluri et al., 2008; Norman-Haignere et al., 2019; Shofner & Chaney, 2013; Walker et al., 2019). One promising future direction may be to use the tasks developed here to search for neural signatures of harmonicity-based sound segregation in noise. It could also be informative to measure the harmonic detection advantage in individuals with listening disorders (Boets et al., 2007; Cameron et al., 2006; Dole et al., 2012; Lagace et al., 2010; Ziegler et al., 2009), as its presence or absence might help pin down the origins of commonly observed hearing-in-noise deficits (Dole et al., 2012; Lagace et al., 2010; Ziegler et al., 2009).

REFERENCES

Assman, P., & Summerfield, Q. (1990). Modeling the perception of concurrent vowels: Vowels with different fundamental frequencies. *Journal of the Acoustical Society of America*, *88*(2), 680-697.

Bacon, S. P., Opie, J. M., & Montoya, D. Y. (1998). The effects of hearing loss and noise masking on the masking release for speech in temporally complex backgrounds. *Journal of Speech and Hearing Research*, *41*, 549-563.

Boebinger, D., Evans, S., Rosen, S., Lima, C. s. F., Manly, T., & Scott, S. K. (2015). Musicians and non-musicians are equally adept at perceiving masked speech. *Journal of the Acoustical Society of America*, *137*, 378-387.

Boets, B., Ghesquière, P., van Wieringen, A., & Wouters, J. (2007). Speech perception in preschoolers at family risk for dyslexia: Relations with low-level auditory processing and phonological ability. *Brain and Language*, *101*, 19-30.

Bregman, A. S. (1990). *Auditory Scene Analysis: The Perceptual Organization of Sound*. MIT Press.

Brunstrom, J. M., & Roberts, B. (2000). Separate mechanisms govern the selection of spectral components for perceptual fusion and for the computation of global pitch. *Journal of the Acoustical Society of America*, *107*(3), 1566-1577.

Cameron, S., Dillon, H., & Newall, P. (2006). The listening in spatialized noise test: An auditory processing disorder study. *Journal of the American Academy of Audiology*, *17*, 306-320.

Carlyon, R. P. (2004). How the brain separates sounds. *Trends in Cognitive Sciences*, *8*(10), 465-471.

Chalikia, M. H., & Bregman, A. S. (1993). The perceptual segregation of simultaneous vowels with harmonic, shifted, or random components. *Perception and Psychophysics*, *53*(2), 125-133.

Clayton, K. K., Swaminathan, J., Yazdanbakhsh, A., Zuk, J., Patel, A. D., & Kidd Jr, G. (2016). Executive function, visual attention and the cocktail party problem in musicians and non-musicians. *PLoS ONE*, *11*(7), e0157638.

Coffey, E. B. J., Mogilever, N. B., & Zatorre, R. J. (2017). Speech-in-noise perception in musicians: A review. *Hearing Research*, *352*, 49-69.

Culling, J. F., & Summerfield, Q. (1995a). Perceptual separation of concurrent speech sounds: Absence of across-frequency grouping by common interaural delay. *Journal of the Acoustical Society of America*, *98*(2), 785-797.

Culling, J. F., & Summerfield, Q. (1995b). The role of frequency modulation in the perceptual segregation of concurrent vowels. *The Journal of the Acoustical Society of America*, *98*(2), 837-846.

Cusack, R., Decks, J., Aikman, G., & Carlyon, R. P. (2004). Effects of location, frequency region, and time course of selective attention on auditory scene analysis. *Journal of Experimental Psychology: Human Perception and Performance*, *30*(4), 643-656.

Darwin, C. J. (1981). Perceptual grouping of speech components differing in fundamental frequency and onset-time. *Quarterly Journal of Experimental Psychology*, *33*(185-207).

Darwin, C. J. (1997). Auditory grouping. *Trends in Cognitive Sciences*, *1*, 327-333.

Darwin, C. J., Brungart, D. S., & Simpson, B. D. (2003). Effects of fundamental frequency and vocal-tract length changes on attention to one of two simultaneous talkers. *Journal of the Acoustical Society of America*, *114*(5), 2913-2922.

Darwin, C. J., & Ciocca, V. (1992). Grouping in pitch perception: effects of onset asynchrony and ear of presentation of a mistuned component. *Journal of the Acoustical Society of America*, *91*, 3381-3390.

Darwin, C. J., & Hukin, R. W. (1997). Perceptual segregation of a harmonic from a vowel by interaural time difference and frequency proximity. *Journal of the Acoustical Society of America*, *102*(4), 2316-2324.

de Cheveigne, A., Kawahara, H., Tsuzaki, M., & Aikawa, K. (1997). Concurrent vowel identification. I. Effects of relative amplitude and F0 difference. *Journal of the Acoustical Society of America*, *101*(5), 2839-2847.

de Cheveigne, A., McAdams, S., Laroche, J., & Rosenberg, M. (1995). Identification of concurrent harmonic and inharmonic vowels: A test of the theory of harmonic cancellation and enhancement. *Journal of the Acoustical Society of America*, *97*(6), 3736-3748.

de Cheveigne, A., McAdams, S., & Marin, C. M. H. (1997). Concurrent vowel identification. II. Effects of phase, harmonicity, and task. *Journal of the Acoustical Society of America*, *101*(5), 2848-2856.

Dellacherie, D., Roy, M., Hugueville, L., Peretz, I., & Samson, S. (2010). The effect of musical experience on emotional self-reports and psychophysiological responses to dissonance. *Psychophysiology*, *48*, 337-349.

Demany, L., & Ramos, C. (2005). On the binding of successive sounds: Perceiving shifts in nonperceived pitches. *Journal of the Acoustical Society of America*, *117*(2), 833-841.

Dole, M., Hoen, M., & Meunier, F. (2012). Speech-in-noise perception deficit in adults with dyslexia: Effects of background type and listening configuration. *Neuropsychologia*, *50*, 1543-1552.

Duquesnoy, A. J. (1983). Effect of a single interfering noise or speech source on the binaural sentence intelligibility of aged persons. *Journal of the Acoustical Society of America*, *74*, 739-743.

Ellis, D. P. W., McDermott, J. H., & Kawahara, H. (2012). Inharmonic Speech: A Tool for the Study of Speech Perception and Separation. *Proc. SAPA - SCALE Conference 2012*, 114-117.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175-191.

Faulkner, A. (1985). Pitch discrimination of harmonic complex signals: Residue pitch or multiple component discriminations? *The Journal of the Acoustical Society of America*, *78*.

Feng, L., & Wang, X. (2017). Harmonic template neurons in primate auditory cortex underlying complex sound processing. *Proceedings of the National Academy of Sciences*, E840-E848.

Festen, J. M., & Plomp, R. (1990). Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing. *Journal of the Acoustical Society of America*, *88*, 1725-1736.

Florentine, M., Buus, S. r., & Bonding, P. (1978). Loudness of complex sounds as a function of the standard stimulus and the number of components. *The Journal of the Acoustical Society of America*, *64*(4), 1036-1040.

Freyman, R. L., Balakrishnan, U., & Helfer, K. S. (2001). Spatial release from informational masking in speech recognition. *Journal of the Acoustical Society of America*, *109*(5), 2112-2122.

Hafter, E. R., & Saberi, K. (2001). A level of stimulus representation model for auditory detection and attention. *The Journal of the Acoustical Society of America*, *110*(3), 1489-1497.

Hall, J. W., & Peters, R. W. (1981). Pitch for nonsimultaneous successive harmonics in quiet and noise. *Journal of the Acoustical Society of America*, *69*(2), 509-513.

Hartmann, W. M., McAdams, S., & Smith, B. K. (1990). Hearing a mistuned harmonic in an otherwise periodic complex tone. *Journal of the Acoustical Society of America*, *88*, 1712-1724.

Hawley, M. L., Litovsky, R. Y., & Culling, J. F. (2004). The benefit of binaural hearing in a cocktail party: Effect of location and type of interferer. *Journal of the Acoustical Society of America*, *115*(2), 833-843.

Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America*, *97*(5), 3099-3111.

Houtgast, T. (1976). Subharmonic pitches of a pure tone at low S/N ratio. *Journal of the Acoustical Society of America*, *60*(2), 405-409.

Ihlefeld, A., & Shinn-Cunningham, B. (2008). Spatial release from energetic and informational masking in a divided speech identification task. *Journal of the Acoustical Society of America*, *123*(6), 4380-4392.

JASP, Version 0.13.1. In. (2020). JASP Team.

Josupeit, A., & Hohmann, V. (2017). Modeling speech localization, talker identification, and word recognition in a multi-talker setting. *The Journal of the Acoustical Society of America*, *142*(1), 35-54.

Josupeit, A., Schoenmaker, E., van de Par, S., & Hohmann, V. (2020). Sparse periodicity-based auditory features explain human performance in a spatial multitalker auditory scene analysis task. *European Journal of Neuroscience*, *51*(5), 1353-1363.

Kalluri, S., Depireux, D. A., & Shamma, S. A. (2008). Perception and cortical neural coding of harmonic fusion in ferrets. *The Journal of the Acoustical Society of America*, *123*(5), 2701-2716.

Kawahara, H., & Morise, M. (2011). TANDEM-STRAIGHT: A temporally stable power spectral representation for periodic signals and applications to interference-free spectrum, F0, and aperiodicity estimation. *SADHANA*, *36*(5), 713-722.

Kell, A. J. E., & McDermott, J. H. (2019). Invariance to background noise as a signature of non-primary auditory cortex. *Nature Communications*, in press.

Khalighinejad, B., Herrero, J. L., Mehta, A. D., & Mesgarani, N. (2019). Adaptation of the human auditory cortex to changing background noise. *Nature Communications*, *10*(1), 2509.

Kishon-Rabin, L., Amir, O., Vexler, Y., & Zaltz, Y. (2001). Pitch discrimination: are professional musicians better than non-musicians? *Journal of Basic Clinical Physiology and Pharmacology*, *12*, 125-143.

Lagace, J., Jutras, B., & Gagne, J. (2010). Auditory processing disorder and speech perception problems in noise: Finding the underlying origin. *American Journal of Audiology*, *19*, 17-25.

Madsen, S., Marschall, M., Dau, T., & Oxenham, A. (2019). Speech perception is similar for musicians and non-musicians across a wide range of conditions. *Scientific Reports*, *9*(1), 1-10.

McAdams, S. (1989). Segregation of concurrent sounds. I.:Effects of frequency modulation coherence. *Journal of the Acoustical Society of America*, *86*, 2148-2159.

McDermott, J. H. (2009). The cocktail party problem. *Current Biology*, *19*, R1024-R1027.

McDermott, J. H., Keebler, M. V., Micheyl, C., & Oxenham, A. J. (2010). Musical intervals and relative pitch: Frequency resolution, not interval resolution, is special. *The Journal of the Acoustical Society of America*, *128*(4), 1943-1951.

McDermott, J. H., Lehr, A. J., & Oxenham, A. J. (2010). Individual differences reveal the basis of consonance. *Current Biology*, *20*, 1035-1041.

McDermott, J. H., & Oxenham, A. J. (2008). Spectral completion of partially masked sounds. *Proceedings of the National Academy of Sciences*, *105*(15), 5939-5944.

McDermott, J. H., Schultz, A. F., Undurraga, E. A., & Godoy, R. A. (2016). Indifference to dissonance in native Amazonians reveals cultural variation in music perception. *Nature*, *535*, 547-550.

McPherson, M. J., Dolan, S. E., Durango, A., Ossandon, T., Valdés, J., Undurraga, E. A., Jacoby, N., Godoy, R. A., & McDermott, J. H. (2020). Perceptual fusion of musical notes by native Amazonians suggests universal representations of musical intervals. *Nature Communications*, *11*(1), 1-14.

McPherson, M. J., & McDermott, J. H. (2018). Diversity in pitch perception revealed by task dependence. *Nature Human Behavior*, *2*, 52-66.

McPherson, M. J., & McDermott, J. H. (2020). Efficient codes for memory determine pitch representations. *bioRxiv*.

McWalter, R., & McDermott, J. H. (2019). Illusory sound texture reveals multi-second statistical completion in auditory scene analysis. *Nature Communications*, *10*(1), 5096.

McWalter, R. I., & McDermott, J. H. (2018). Adaptive and selective time-averaging of auditory scenes. *Current Biology*, *28*, 1405-1418.

Mesgarani, N., David, S. V., Fritz, J. B., & Shamma, S. A. (2014). Mechanisms of noise robust representation of speech in primary auditory cortex. *Proceedings of the National Academy of Sciences*, *111*(18), 6792-6797.

Micheyl, C., Delhommeau, K., Perrot, X., & Oxenham, A. J. (2006). Influence of musical and psychoacoustical training on pitch discrimination. *Hearing Research*, *219*, 36-47.

Micheyl, C., & Oxenham, A. J. (2010). Pitch, harmonicity and concurrent sound segregation: Psychoacoustical and neurophysiological findings. *Hearing Research*, *266*, 36-51.

Micheyl, C., Ryan, C. M., & Oxenham, A. J. (2012). Further evidence that fundamental-frequency difference limens measure pitch discrimination. *The Journal of the Acoustical Society of America*, *131*(5), 3989-4001.

Młynarski, W., & McDermott, J. H. (2019). Ecological origins of perceptual grouping principles in the auditory system. *Proceedings of the National Academy of Sciences*, *116*(50), 25355-25364.

Moore, B. C., Peters, R. W., & Glasberg, B. R. (1985). Thresholds for the detection of inharmonicity in complex tones. *The Journal of the Acoustical Society of America*, *77*(5), 1861-1867.

Moore, B. C. J. (1987). The perception of inharmonic complex tones. In W. A. Yost & C. S. Watson (Eds.), *Auditory Processing of Complex Sounds* (pp. 180-189). Erlbaum.

Moore, B. C. J., & Glasberg, B. R. (1990). Frequency discrimination of complex tones with overlapping and non-overlapping harmonics. *The Journal of the Acoustical Society of America*, *87*.

Moore, B. C. J., Glasberg, B. R., & Peters, R. W. (1986). Thresholds for hearing mistuned partials as separate tones in harmonic complexes. *Journal of the Acoustical Society of America*, *80*, 479-483.

Moore, B. C. J., Huss, M., Vickers, D. A., Glasberg, B. R., & Alcántara, J. I. (2000). A test for the diagnosis of dead regions in the cochlea. *British Journal of Audiology*, *34*(4), 205-224.

Moore, R. C., Lee, T., & Theunissen, F. E. (2013). Noise-invariant neurons in the avian auditory cortex: hearing the song in noise. *PLoS computational biology*, *9*(3), e1002942.

Norman-Haignere, S., & McDermott, J. H. (2016). Distortion products in auditory fMRI research: Measurements and solutions. *Neuroimage*, *in press*.

Norman-Haignere, S. V., Kanwisher, N. G., McDermott, J. H., & Conway, B. R. (2019). Divergence in the functional organization of human and macaque auditory cortex revealed by fMRI responses to harmonic tones. *Nature Neuroscience*, *22*, 1057-1060.

Oxenham, A. J. (2008). Pitch perception and auditory stream segregation: Implications for hearing loss and cochlear implants. *Trends in Amplification*, *12*(4), 316-331.

Oxenham, A. J., Fligor, B. J., Mason, C. R., & Kidd, G. (2003). Informational masking and musical training. *Journal of the Acoustical Society of America*, *114*(3), 1543-1549.

Parbery-Clark, A., Strait, D. L., Anderson, S., Hittner, E., & Kraus, N. (2011). Musical experience and the aging auditory system: implications for cognitive abilities and hearing speech in noise. *PLoS ONE*, *6*(5), e18082.

Plack, C. J., Barker, D., & Prendergast, G. (2014). Perceptual consequences of "hidden" hearing loss. *Trends in Hearing*, *18*, 1-11.

Popham, S., Boebinger, D., Ellis, D. P., Kawahara, H., & McDermott, J. H. (2018). Inharmonic speech reveals the role of harmonicity in the cocktail party problem. *Nature Communications*, *9*(1), 2122.

Pressnitzer, D., & Patterson, R. D. (2001). *Distortion products and the perceived pitch of harmonic complex tones*. http://lumiere.ens.fr/Audition/dp/pdfs/pressnitzer-2001-distortion_pitch.pdf

Rabinowitz, N. C., Willmore, B. D. B., King, A. J., & Schnupp, J. W. (2013). Constructing noise-invariant representations of sound in the auditory pathway. *PLoS Biology*, *11*(11), e1001710.

Rasch, R. A. (1978). The perception of simultaneous notes such as in polyphonic music. *Acustica*, *40*, 21-33.

Roberts, B., & Brunstrom, J. M. (1998). Perceptual segregation and pitch shifts of mistuned components in harmonic complexes and in regular inharmonic complexes. *Journal of the Acoustical Society of America*, *104*(4), 2326-2338.

Roberts, B., & Brunstrom, J. M. (2001). Perceptual fusion and fragmentation of complex tones made inharmonic by applying different degrees of frequency shift and spectral stretch. *Journal of the Acoustical Society of America*, *110*(5), 2479-2490.

Roberts, B., & Holmes, S. D. (2006). Grouping and the pitch of a mistuned fundamental component: effects of applying simultaneous multiple mistunings to the other harmonics. *Hearing Research*, *222*, 79-88.

Rossi-Katz, J. A., & Arehart, K. H. (2005). Effects of cochlear hearing loss on perceptual grouping cues in competing-vowel perception. *Journal of the Acoustical Society of America*, *118*(4), 2588-2598.

Ruggles, D., Bharadwaj, H., & Shinn-Cunningham, B. G. (2012). Why middle-aged listeners have trouble hearing in everyday settings. *Current Biology*, *22*(15), 1417-1422.

Sarampalis, A., Kalluri, S., Edwards, B., & Hafter, E. (2009). Objective measures of listening effort: Effects of background noise and noise reduction. *Journal of Speech, Language, and Hearing Research*.

Shackleton, T. M., & Carlyon, R. P. (1994). The role of resolved and unresolved harmonics in pitch perception and frequency modulation discrimination. *Journal of the Acoustical Society of America*, *95*(6), 3529-3540.

Shofner, W. P., & Chaney, M. (2013). Processing pitch in a nonhuman mammal (chinchilla laniger). *Journal of Comparative Psychology*, *127*(2), 142-153.

Smoorenburg, G. F. (1992). Speech reception in quiet and in noisy conditions by individuals with noise-induced hearing loss in relation to their tone audiogram. *The Journal of the Acoustical Society of America*, *91*.

Spiegel, M. F., & Watson, C. S. (1984). Performance on frequency-discrimination tasks by musicians and non-musicians. *Journal of the Acoustical Society of America*, *76*, 1690-1695.

Swaminathan, J., Mason, C. R., Streeter, T. M., Best, V., Kidd, G., & Patel, A. (2015). Musical training, individual differences and the cocktail party problem. *Scientific Reports*, *5*, 11628.

Town, S. M., Wood, K. C., & Bizley, J. K. (2019). Signal processing in auditory cortex underlies degraded speech sound discrimination in noise. *bioRxiv*, 833558.

Tremblay, K. L., Piskosz, M., & Souza, P. (2003). Effects of age and age-related hearing loss on the neural representation of speech cues. *Clinical Neurophysiology*, *114*(7), 1332-1343.

Walker, K. M. M., Gonzalez, R., Kang, J. Z., McDermott, J. H., & King, A. J. (2019). Across-species differences in pitch perception are consistent with differences in cochlear filtering. *eLIFE*, *8*, e41626.

Weiss, M. W., Cirelli, L. K., McDermott, J. H., & Trehub, S. E. (2019). Development of consonance preferences in Western listeners. *Journal of experimental psychology. General*.

Woods, K. J. P., & McDermott, J. (2018). Schema learning for the cocktail party problem. *Proceedings of the National Academy of Sciences*, *115*, E3313-E3322.

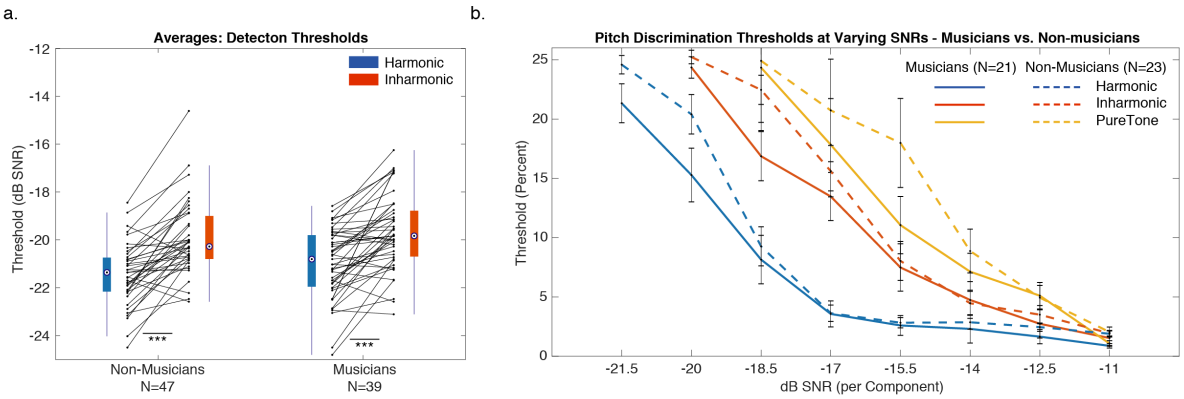
Woods, K. J. P., & McDermott, J. H. (2015). Attentive tracking of sound sources. *Current Biology*, *25*, 2238-2246.

Woods, K. J. P., Siegel, M. H., Traer, J., & McDermott, J. H. (2017). Headphone screening to facilitate web-based auditory experiments. *Attention, Perception, and Psychophysics*, *79*, 2064-2072.

Zendel, B. R., & Alain, C. (2009). Concurrent sound segregation is enhanced in musicians. *Journal of Cognitive Neuroscience*, *21*(8), 1488-1498.

Ziegler, J. C., Pech-Georgel, C., George, F., & Lorenzi, C. (2009). Speech-perception-in-noise deficits in dyslexia. *Developmental Science*, *12*(5), 732-745.

Zwicker, E. (1965). Temporal effects in simultaneous masking and loudness. *The Journal of the Acoustical Society of America*, *38*(1), 132-141.



Supplementary Figure 1: Effects of Musicianship on Detection and Discrimination in Noise
a. Results of Experiment 1, separated for musicians and non-musicians. The results are averaged across sine and random phase conditions, and shown as box-and-whisker-plots, with black lines plotting individual participant results. The central mark in the box plots the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. Whiskers extend to the most extreme data points not considered outliers. ***=p<0.001. b. Results from Experiment 2, separated for musicians and non-musicians.