Exploring the impact of perceived loudness on a preference toward spectral or virtual pitch listening in tone-pair motion discrimination

Emma McGonigle[1,2], Tyler Furrier[1,3], Maximilian Bateman[1], Jacob Allen[1,4]

Northeastern University
April 16th, 2024
Music and the Brain Research
Dr. Psyche Loui

1. Northeastern University College of Arts, Media and Design, Department of Music
2. Northeastern University College of Science, Department of Psychology
3. Northeastern University Khoury College of Computer Science
4. Northeastern University College of Science, Department of Physics

Abstract:

The perception of harmonically complex tones can occur in multiple distinct ways. While the mind can primarily detect the pitch associated with a harmonic partial with a dramatic power level (listening "spectrally"), it can also perceive a lower fundamental pitch that is masked or not present, this pitch being determined by the other harmonic partials in the series (listening "virtually"). Within the auditory cortex, a significant amount of neurons (~20% in similar primates) process harmonic information, as well as collections of neurons dedicated to individual ranges of pitches, similar to how ear hair cells are "tuned" to more specific frequencies. It's likely that the mind does not perceive sounds categorically as exclusively spectral or virtual information, but rather does so on a sliding scale. The following study aimed to disprove a bimodal distribution of "spectral" and "virtual" pitch perception by observing and statistically analyzing how volume affects where listeners skew along the scale of spectral to virtual.

Introduction:

All complex sounds can be broken down, mathematically, into a sequence of sinusoidal components (Schneider, 2018). Pitch rises from the perceptual correlates of following the periodicity of said sinusoidal components of the waveform. Most perceived pitches are not pure sinusoidal functions, but rather a complex harmonic series consisting of a fundamental frequency (f0, in hz), and its harmonic series (2*f0, 3*f0, 4*f0, and so on)(Oxenham, 2012). The fundamental frequency is the common denominator of these harmonic frequencies and the convergence point of the periodic sinusoidal waveforms which create the harmonic series. In some circumstances, the brain will perceive a pitch that isn't actually in a tone but simply generated in the mind as a result of hearing the cumulative effect of many harmonic partials (Dai, 2010). In reality, the psychoacoustic effect of pitch, as a whole, is the composite perception of a fusion of fundamental frequencies and harmonic partial information (Schneider, 2018). This

suggests that the brain tends to homogenize together harmonically related data (Micheyl et al., 2010). The conglomerate sound is referred to as “virtual pitch”. Even when the fundamental pitch is missing from the tone series, neural correlates suggest similar responses to harmonics with a common fundamental frequency (Zatorre, 2005). One can hear a sound “spectrally” (wherein they hear the pitch associated with a particular partial(s) in the sound), or “virtually” (in which a fundamental frequency that isn’t present is perceived audibly by the listener).

The reward pathway lies mostly in the midbrain and prefrontal cortex (Chau et al., 2018). Auditory processing and categorization seem to lie in the superior temporal gyrus of the temporal lobe, though some information is also received through the medial geniculate complex of the midbrain (Steinschneider, 2011) (Winer, 1984). Dopaminergic pathways, in essence, are responsible for movement, control, executive function, reward, and motivation, among other important behaviors (Gepshtein et al., 2014)(Floresco et al., 2006)(Nieoullon et al., 2003). If an individual has low reward sensitivity to music, they are less likely to have a high dopaminergic response to music, which will uniquely shape their neural mapping (Belfi et al., 2020). The saying “neurons that fire together, wire together” is an example of how neuronal response, on a basic level, can change our neuronal structures and pathways. It has been hypothesized that neuronal asymmetry may be a factor in the preference of spectral (Fsp) or fundamental (F0) pitch perception (Schneider et al., 2002). It is possible that the utilization, or neglect of certain dopaminergic responses to music could have an impact on the structural construction of asymmetrical areas of the brain.

While the human pitch center of the brain doesn’t lie exclusively in one area, there is markedly more neural activity in the lateral Heschl’s gyrus and the planum temporale. The Heschl’s Gyrus in particular is the area stimulated first and most frequently by pitch modulation

(Yuskaitis et al, 2017). The left hemisphere tends to specialize in processing temporal auditory data, whereas the left corresponds more to the perception of spectral harmonic simulation. (Zatorre and Belin, 2001). Hair cells within the cochlea and cells within the auditory nerve are each tuned to specific frequencies (Kiang, 1967)(Schneider, 2018). However, it has been found that within the auditory cortex, a significant portion of the neurons are dedicated to processing harmonic information. They can do multi-peak frequency tuning, which entails the deduction of a virtual pitch from the cumulative effect of multiple harmonic partials (Marsh et. al. 2006).

It is feasible to quantify this phenomenon algorithmically using the autocorrelation function (ACF) on neurological data. The ACF of the nerve firing patterns in eight fibers was taken. When presented with sounds with high harmonic complexity, there were peaks corresponding to integer multiples of the period of the stimulus (Meddis and Hewitt, 1991). If two such sounds were played in tandem, the ACFs had peaks that reflected both fundamental frequencies (assuming they aren't close together or in an octave relationship)(Meddis and Hewitt, 1991). With the aforementioned conditions present, it is possible to use the ACF algorithm to determine multiple fundamental frequencies by "peak picking" using exclusively psychoelectric information.

It is yet to be determined whether this is a binary selection between the two categories, or whether the mind can perceive this difference along a spectrum (Ladd et al., 2013). It should be noted that a correlation between musical training and increased ability for pitch discrimination has been found in some studies, which may have an overall influence on spectral versus virtual pitch listening, contributing to the "sliding scale" of pitch perception (Micheyl et al., 2012). Both spectral and virtual pitch perception play a crucial role in pitch discrimination, and some listeners may be able to interchangeably distinguish spectrally, or by isolating the fundamental in

later stages of auditory processing (Coffey et al., 2016). Many factors go into this distinction, more of which are yet to be discovered. One factor that has not, to our knowledge, been investigated yet is the role of perceived loudness in spectral versus virtual listening. In this study, we hypothesize that the perceived loudness of a sound will influence whether a listener discerns a pitch change by listening spectrally, or virtually.

Methods:

Participants:

Participants were recruited through PsyLink, as well as social media and student body populations at Northeastern University. Participants self-reported normal hearing, no hearing loss, and had access to a main device that played audio. This sample had 38 participants.

Stimuli:

Research stimuli consisted of several pairs of complex harmonic tones with the desired percept of moving either up or down depending on the listening mode (spectral or virtual). Pertinently, the perceived loudness of the tones needed to be isolated. An approach similar to a previous study was used to determine this (Ladd et al., 2013). Each tone in the pair consists of three upper partials. The highest partial will always be the same in both tones, facilitating the phenomenon of the spectrum moving in the opposite direction of the movement of the fundamental (Ladd et al., 2013). Figure 1 shows the tone pairs and their respective configurations.

In order to isolate and associate a value of perceived loudness to each tone pair, ISO 226:2023[1] standards were applied to normalize the pure tone components of the tones to a perceived loudness. Approximate amplitude/voltage to physical sound pressure conversions were deduced by assuming a flat frequency response, and a linear relationship between voltage and

pressure on the listener's apparatus. A tone that, in calibration with the SPL's of the stimuli, would be exactly the hearing threshold[1] was used for the participant to calibrate their apparatus. STC: Experimentation on various listening devices available influenced the reference amplitude, aiming to keep hardware volume levels as low as possible to avoid uncontrollable hardware clipping or faults in our previous assumption of linear voltage-to-pressure relationships. We chose to perform linear interpolation, for undefined frequencies, out of ease and replicability. The likelihood of our results being dependent on the form of interpolation chosen is presumed to be extremely low. The authors, all musically trained, found the tones to be equally loud. Moreover, the effect of the various interpolations is marginal due to the dense sampling in ISO studies. This math allows for a more universal measurement that allows for replicability and potential compatibility with other research. Importantly for this study, controlling for loudness allows for tone pairs that use various tone configurations while controlling for the perceived loudness. For each tone pair, there's a quiet, medium, and loud iteration, at 10, 30 and 50 phon respectively. It is worth noting that each component of the tone is normalized to that loudness, so a 10 phon tone is realistically more than 10 phon by some amount depending on the interaction of the tones. However, controlling the loudness of the individual frequencies best tailors the stimuli to the research question.

A test tone of 1000 hz was used for calibration to the participants hearing threshold. In doing so, a relatively consistent digital amplitude (voltage) to pressure ratio for the participant's listening apparatus was developed. Participants were instructed to adjust the volume to the hearing threshold[1] of the test tone, which is 2.4 db SPL for 1,000hz[1]. STC: As seen on the associated GitHub repository[2], an arbitrary[2] level was assigned to 80dBSPL at an impossible

[1] Defined in ISO as being detected 50% of the time
[2] This level is not completely arbitrary. Increasing it will increase the volume of the test tone and allow the user to set their hardware level lower.

voltage of 2 (max being 1). This is then used to produce a pure tone at a particular volume. The congruency between participants' volume level assumes a perfectly flat frequency response, as the ratio of voltage (amplitude) to pressure may differ in frequency ranges differing from 1000 hz. This method also disregards any potential clipping that occurs as a result of hardware, which would have a distortion effect on harmonic partials.

For each tone pair, a "quiet" set was developed where each harmonic component was at a level roughly perceived at 10 phon. The same approach was used to create "medium" and "loud" sets where components are normalized to STC:30 and 50 phon respectively. Notably, tones composed of harmonic partials which are closer in frequency combined to be perceived slightly louder than tones with a greater frequency spread. This effect was most pronounced for tones in the 10+kHz range, possibly because they lie within the same band according to critical band theory. STC: Investigation into the inter-harmonic distances within a tone and across two tones in a pair is still opportunistic to better understand and predict the behavior of our stimuli configurations.

Low loudness	Mid loudness	High loudness
Set 1 (pitch 1- pitch 2)	Set 1 (pitch 1- pitch 2)	Set 1 (pitch 1- pitch 2)
Set 2 (pitch 3 - pitch 4)	Set 2 (pitch 3 - pitch 4)	Set 2 (pitch 3 - pitch 4)
Set 3 (pitch 5 - pitch 6)	Set 3 (pitch 5 - pitch 6)	Set 3 (pitch 5 - pitch 6)
Set 4 (pitch 7 - pitch 8)	Set 4 (pitch 7 - pitch 8)	Set 4 (pitch 7 - pitch 8)
Set 5 (pitch 9 -pitch 10)	Set 5 (pitch 9 -pitch 10)	Set 5 (pitch 9 -pitch 10)
Set 6 (pitch 11 - pitch 12)	Set 6 (pitch 11 - pitch 12)	Set 6 (pitch 11 - pitch 12)
Set 7 (pitch 13 - pitch 14)	Set 7 (pitch 13 - pitch 14)	Set 7 (pitch 13 - pitch 14)
Set 8 (pitch 15 - pitch 16)	Set 8 (pitch 15 - pitch 16)	Set 8 (pitch 15 - pitch 16)
Set 9 (pitch 17 -pitch 18)	Set 9 (pitch 17 -pitch 18)	Set 9 (pitch 17 -pitch 18)
Set 10 (pitch 19 - pitch 20)	Set 10 (pitch 19 - pitch 20)	Set 10 (pitch 19 - pitch 20)

Procedure:

A Qualtrics survey was developed to determine how volume affects the perception of a fundamental frequency f0 in a complex tone. Preceding the trials, participants adjusted the loudness level of a 1000 hz test tone to a point at which it was barely audible. The purpose of this test tone was to calibrate the sound pressure levels of the participant's listening device to the digital amplitude of the stimuli. The survey began with a set of STC: ten trials conducted using two pitches each. Each set of these pitches was equalized to the same level of loudness according to the IS curve to dB shift. These two tones played in sequence at a low, medium, and high loudness level, respectively. Each time the two tones played, the user was asked whether the pitch rose, fell, or remained the same between the two. This trial was then repeated with ten more pairs of tones (for a total of 12). Out of this set of tests, two pitch pairs served as a catch trial, functioning as screening tones where the pitch decreases both spectrally and virtually. This catch-trial facilitates the elimination of data from participants who are not paying attention to the study as they take it in addition to participants who could have been tone-deaf. The study then utilizes a combination of within-subject and between-subject comparisons.

Data Analysis:

Participants included in this study completed the full survey and had no known hearing loss or congenital amusia. Exclusion criteria included participants who incorrectly answered *both* screening tones. We hypothesized that differences in spectral versus virtual pitch listening are dependent on the perceived loudness of a given sample. To test this hypothesis, we used a within-subject repeated measures ANOVA using SPSS. Our two factors were loudness (3 levels: 10 phon, 30 phon, and 50 phon) and stimuli (10 different two-tone pairs, with an increase in

harmonic spectral frequencies, and a decrease in the virtual fundamental frequency). We assigned response values 1= “up,” 2= ”down,” 3= ”both,” and 4= ”neither,” in regards to if the participant heard the pitch go up, or down (listening spectrally, or virtually respectively).

Results:

The mean response for each loudness level, independent of stimuli, fell within a range of 2.109-1.97, with a standard error between 0.067 and 0.082 (Table 1). The mean response for each stimulus, independent of loudness level, fell between 1.980 and 2.696, with a standard error between 0.077 and 2.369 (Table 2). The mean responses for each independent stimulus, at each loudness level, ranged from 1.68-2.82, with a standard deviation of 0.535-1.258 (Table 3). Marginal significance was found in the above-mentioned repeated measures ANOVA for within-factor analysis of effect of loudness on response ($F(2,35) = 2.930$, $p = 0.60$, partial eta squared = 0.082), and a significant effect of stimulus on response was found ($F(9,25) = 5.822$, $p = <.001$, partial eta squared = 0.150). There was no

Estimates

Measure: response

loudness	Mean	Std. Error	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
1	2.297	.082	2.130	2.465
2	2.141	.072	1.995	2.288
3	2.109	.067	1.972	2.246

Table 1

Descriptive Statistics

	Mean	Std. Deviation	N
1_10_phon	2.09	.900	34
2_10_phon	2.09	.712	34
3_10_phon	2.15	.657	34
4_10_phon	2.32	1.065	34
6_10_phon	2.59	1.209	34
8_10_phon	2.82	1.218	34
22_10_phon	2.21	.880	34
23_10_phon	2.12	.913	34
24_10_phon	2.50	1.135	34
26_10_phon	2.09	.933	34
1_30_phon	2.00	.853	34
2_30_phon	2.21	.914	34
3_30_phon	1.91	.668	34
4_30_phon	2.18	1.086	34
6_30_phon	2.41	1.258	34
8_30_phon	2.65	1.178	34
22_30_phon	1.97	.870	34
23_30_phon	2.09	.866	34
24_30_phon	2.12	.946	34
26_30_phon	1.88	.844	34
1_50_phon	1.68	.535	34
2_50_phon	1.91	.793	34
3_50_phon	2.09	.668	34
4_50_phon	2.06	.983	34
6_50_phon	2.44	1.078	34
8_50_phon	2.62	1.155	34
22_50_phon	2.09	.712	34
23_50_phon	2.15	.925	34
24_50_phon	2.09	.830	34
26_50_phon	1.97	.758	34

Table 3

Estimates

Measure: response

stimulus	Mean	Std. Error	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
1	1.922	.092	1.734	2.109
2	2.069	.087	1.893	2.245
3	2.049	.069	1.908	2.190
4	2.186	.116	1.950	2.423
5	2.480	.149	2.177	2.784
6	2.696	.161	2.369	3.023
7	2.088	.077	1.931	2.246
8	2.118	.100	1.913	2.322
9	2.235	.109	2.013	2.458
10	1.980	.094	1.788	2.172

Table 2

significant effect of loudness on stimulus-response found (F(18,875)=.574, p=0.919, partial eta squared=0.017).

When excluding 3 (both) and 4 (neither) responses, the mean response for each loudness level, independent of stimuli, fell within a range of 1.6-1.8, with a standard error between 0.5 and 0.4 (Table 4). The mean response for each stimulus, independent of loudness level, fell between 1.333 and 2, with a standard error between 0 and 0.5 (Table 5). The mean responses for each independent stimulus, at each loudness level, ranged from 1 to 2, with a standard deviation of 0 to .707 (Table 6). No significance was found in the above-mentioned repeated measures ANOVA for within-factor analysis of the effect of loudness on response (F(2,35) = .302, p = 0.768, partial eta squared = 0.674), and a significant effect of stimulus on response (F(9,25) = 0.302, p = 0.768, partial eta squared = 0.232). There was a marginally significant effect found of loudness on stimulus-response found (F(18,875)=2.064, p=0.067, partial eta squared=0.674).

Estimates

Measure: MEASURE_1

loudness	Mean	Std. Error	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
1	1.750	.050	1.115	2.385
2	1.600	.400	-3.482	6.682
3	1.800	.100	.529	3.071

Table 4

Estimates

Measure: MEASURE_1

stimuli	Mean	Std. Error	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
1	1.667	.000	1.667	1.667
2	2.000	.000	2.000	2.000
3	1.833	.167	-.284	3.951
4	1.667	.333	-2.569	5.902
5	1.333	.333	-2.902	5.569
6	1.500	.500	-4.853	7.853
7	1.833	.167	-.284	3.951
8	1.833	.167	-.284	3.951
9	1.667	.000	1.667	1.667
10	1.833	.167	-.284	3.951

Table 5

Descriptive Statistics

	Mean	Std. Deviation	N
1_10_phon	2.00	.000	2
1_30_phon	2.00	.000	2
1_50_phon	1.00	.000	2
2_10_phon	2.00	.000	2
2_30_phon	2.00	.000	2
2_50_phon	2.00	.000	2
3_10_phon	2.00	.000	2
3_30_phon	1.50	.707	2
3_50_phon	2.00	.000	2
4_10_phon	1.50	.707	2
4_30_phon	1.50	.707	2
4_50_phon	2.00	.000	2
6_10_phon	1.00	.000	2
6_30_phon	1.50	.707	2
6_50_phon	1.50	.707	2
8_10_phon	1.50	.707	2
8_30_phon	1.50	.707	2
8_50_phon	1.50	.707	2
22_10_phon	2.00	.000	2
22_30_phon	1.50	.707	2
22_50_phon	2.00	.000	2
23_10_phon	2.00	.000	2
23_30_phon	1.50	.707	2
23_50_phon	2.00	.000	2
24_10_phon	1.50	.707	2
24_30_phon	1.50	.707	2
24_50_phon	2.00	.000	2
26_10_phon	2.00	.000	2
26_30_phon	1.50	.707	2
26_50_phon	2.00	.000	2

Table 6

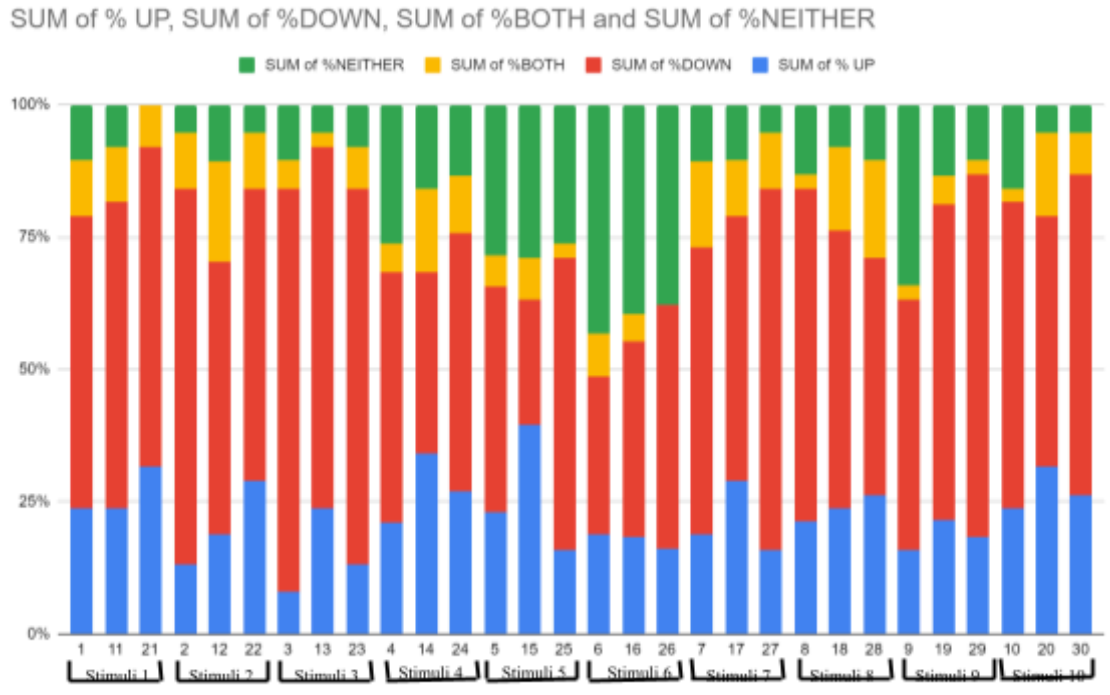
Discussion:

Pitch is a complex perceptual phenomenon, which translates sinusoidal stimuli to create harmonic components by interpreting their corresponding periodicities (Schneider, 2018). There are two well-established means of hearing, or interpreting these pitches, and their changes: either by listening spectrally (hearing the full spectrum of the sinusoidal frequency components) or virtually (interpreting the periodicity of the harmonic series to discern a fundamental, or common denominator frequency, even in the absence of said frequency in the stimulus). Neuronal asymmetry may be an important factor in individual preferences or tendencies to listen spectrally or virtually, with markedly higher activity in the lateral Heschl's gyrus and planum temporale for interpreting pitch modulation (Schneider et al., 2002)(Yuskaitis et al., 2017). Many factors influence this preference for virtual versus spectral listening, including musical training, and some have even asserted that there isn't an either-or scenario at play, but rather a sliding-scale of preferences (Mecheyl et al., 2012)(Ladd et al., 2013). We studied the role of perceived loudness in preference for spectral versus virtual pitch listening.

To test viability we initiated a pilot run of our study. This provided crucial feedback on our methods, but most of all, on the stimuli themselves. The original tone-pairs were created using a Max MSP patch and aimed to exploit different spectral ranges of the same change in fundamental frequency. Besides introducing a potential issue of priming our participants, it also added a fun new problem: when you compare higher-order harmonics (>100*f0), it becomes very difficult to notice small changes in frequency. Frequency is on a logarithmic scale so a change of ~100Hz becomes nearly impossible to perceive when you approach the 20kHz limit, thus one would hear no change in pitch. We received further feedback indicating participants couldn't decide whether the "easier" tone pairs were going up or down, sometimes hearing no

change, and sometimes hearing “both”. As a result, we decided to add options for “neither” and “both” to see if the data would reveal a novel way of viewing the study. In the addition and recreation of tone pairs, they were more evenly spaced out across the 20Hz-20kHz range with appropriate, perceivable harmonics (or at least none that we deemed as having “no change”). Finally, using G*Power we found our new study would require 30 participants to have 80% confidence using the standards of Cohen's f = 0.25.

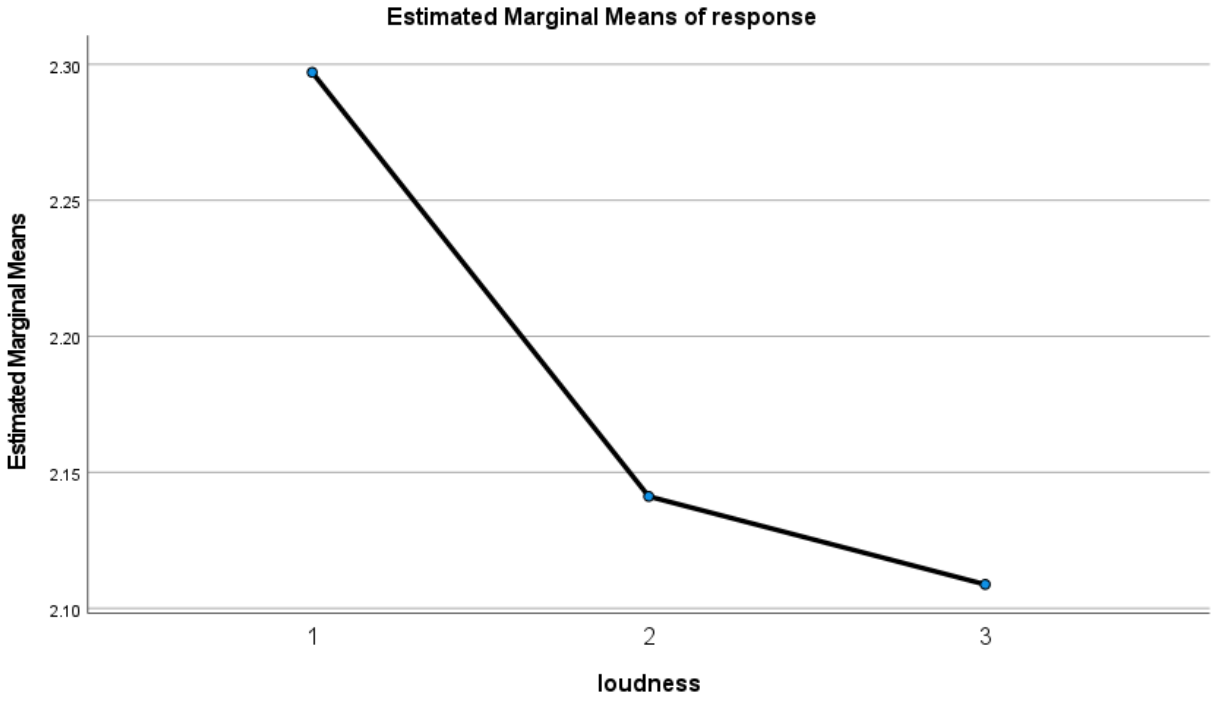
In the finalized version of the study, marginal significance was found for within-factor analysis of effect of loudness on response ($F(2,35) = 2.930$, $p = 0.60$, partial eta squared = 0.082). A significant effect of stimulus on response was found ($F(9,25) = 5.822$, $p = <.001$, partial eta squared = 0.150). There was no significant effect of loudness on stimulus-response found ($F(18,875) = .574$, $p = 0.919$, partial eta squared = 0.017). Overall, our results are in line with previous studies on this topic that yield an overwhelming bias towards f0 “virtual responses” as most test tones favored these responses independent of loudness. The figure below shows the distribution of responses for each stimulus. The red bars represent the percentage of listeners



Percentage of Responses by Stimuli

who heard the pitch move “down,” and were therefore listening virtually. Loudness, however, is where we sought to shake things up, hypothesizing that it may play a factor in favoring “spectral” over “virtual” listening.

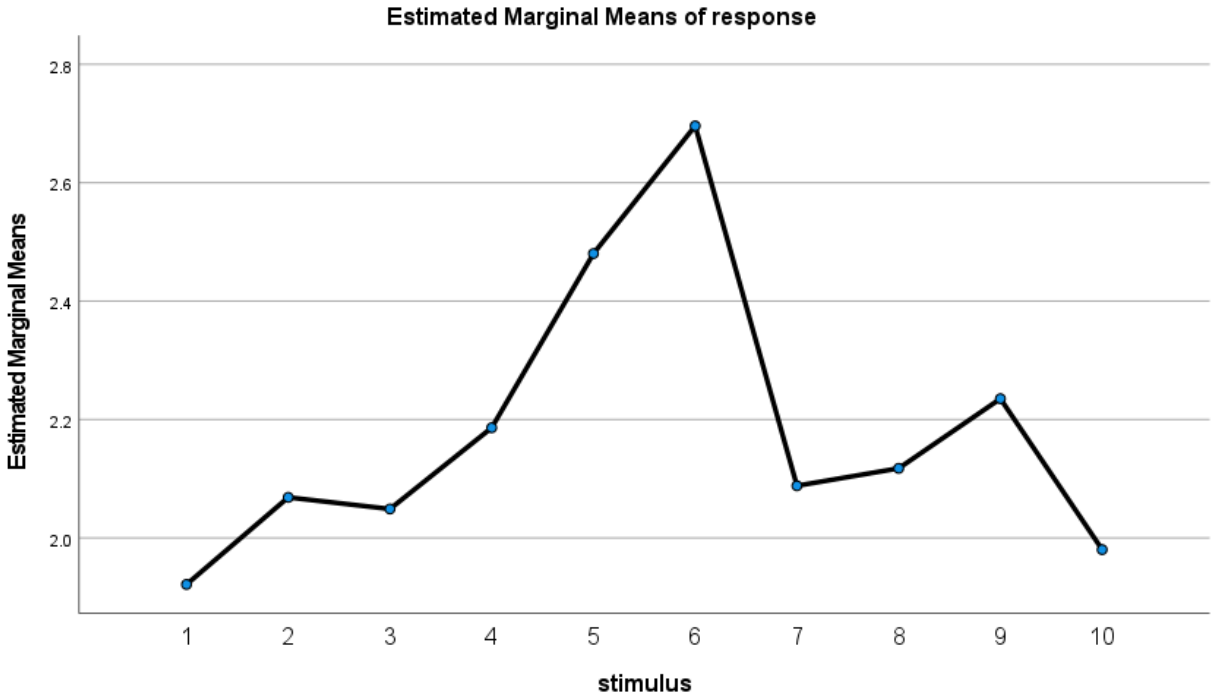
However, our result isn’t crystal clear. When removing “both” or “neither” each loudness level, independent of the stimulus, averaged to 1.6-1.8, indicating a slight bias towards hearing the pitch drop in the absence of a fundamental (with 1.5 being the “middle” of the scale). When looking at our three loudness levels plotted against estimated marginal means (plotted below), it



is evident that an increase in loudness tends to sway responses towards spectral/virtual and away from both/neither, as the estimated marginal mean decreases with an increase in perceived loudness. It still fails to show a clear bias for spectral listening, though, as its averages still favor virtual listening.

Our study does, however, build on the claims that there is no bimodal distribution of "spectral" and "virtual" listeners among us, but rather factors that will favor listeners one way or the other on a spectrum (Ladd et al., 2013). We agree these categories are an oversimplification of the phenomenon, and allow for choices like "neither" and "both" to better understand and diagnose circumstances where an intent listener may fail to hear a designated spectral shift, or where one could switch between methods of analysis due to various qualities of a tone pair. Relevant stimuli factors and various hypotheses of their effects are explored further in the exploratory analysis.

With this frame of mind, we intended to further investigate, rather than exclude, results where participants did not favor spectral or virtual listening. Seither-Preisler concluded a quarter of their participants in this trend were "guessing", excluding them from further analysis. Our inclusion of the "red herring" screening tones (two fundamentals, traveling down a perfect fifth, or up a minor second) intended to remove the possibility, indicating that all respondents have at least a fair sense of relative pitch. Even within the bank of stimuli, some stimuli, regardless of



loudness, were harder to discern than other tones at the same level of loudness, as demonstrated above by the estimated marginal means of response by stimuli. Stimuli 6 seemed to be harder to discern, with the marginal means peaking around that stimulus across loudness level

Exploratory Analysis

By allowing for answers of "both" and "neither", we discovered unexplained perception. There was presumed to be variability in attributes of the stimuli that could be cause for differences in the proportion of unanticipated responses. In attempts to diagnose the causes for these answers, tone pairs were retroactively annotated with features like harmonic rank of the highest partial, spectral centroid shift, and inter-harmonic distances. Inter-harmonic and adjacent-harmonic distances were computed in an attempt to emulate a novel mode of listening in which a participant could be tracking individual partials. The inter-harmonic distance could be 1 to 3, and adjacent-harmonic distance could only be 3 to 2 or 2 to 3. Also, these values are simply off-the-cuff heuristics and most likely to not encapsulate the complex cognitive mechanisms used to align shifted tones.

In theory, if the rate of neither/both responses are accounted for by these new metrics, it might suggest the possibility of unexpected modes of listening. For the needs of this type of study, the complex cognitive process of tracking related harmonics might not be adequately represented by spectral centroid and virtual pitch alone. If ambiguous responses are accounted for by *inter-harmonic*, then there may even be an ability for the human brain to selectively mute tones that do not align with the perceived motion. Ultimately, our logic for computing these jumps between partials ended up returning values as high as half steps for

max-adjacent-harmonic distance in tone pairs with the smallest spectral centroid shift. Thus, the metrics were irrelevant, more likely due to the algorithm[3] than a lack of significance.

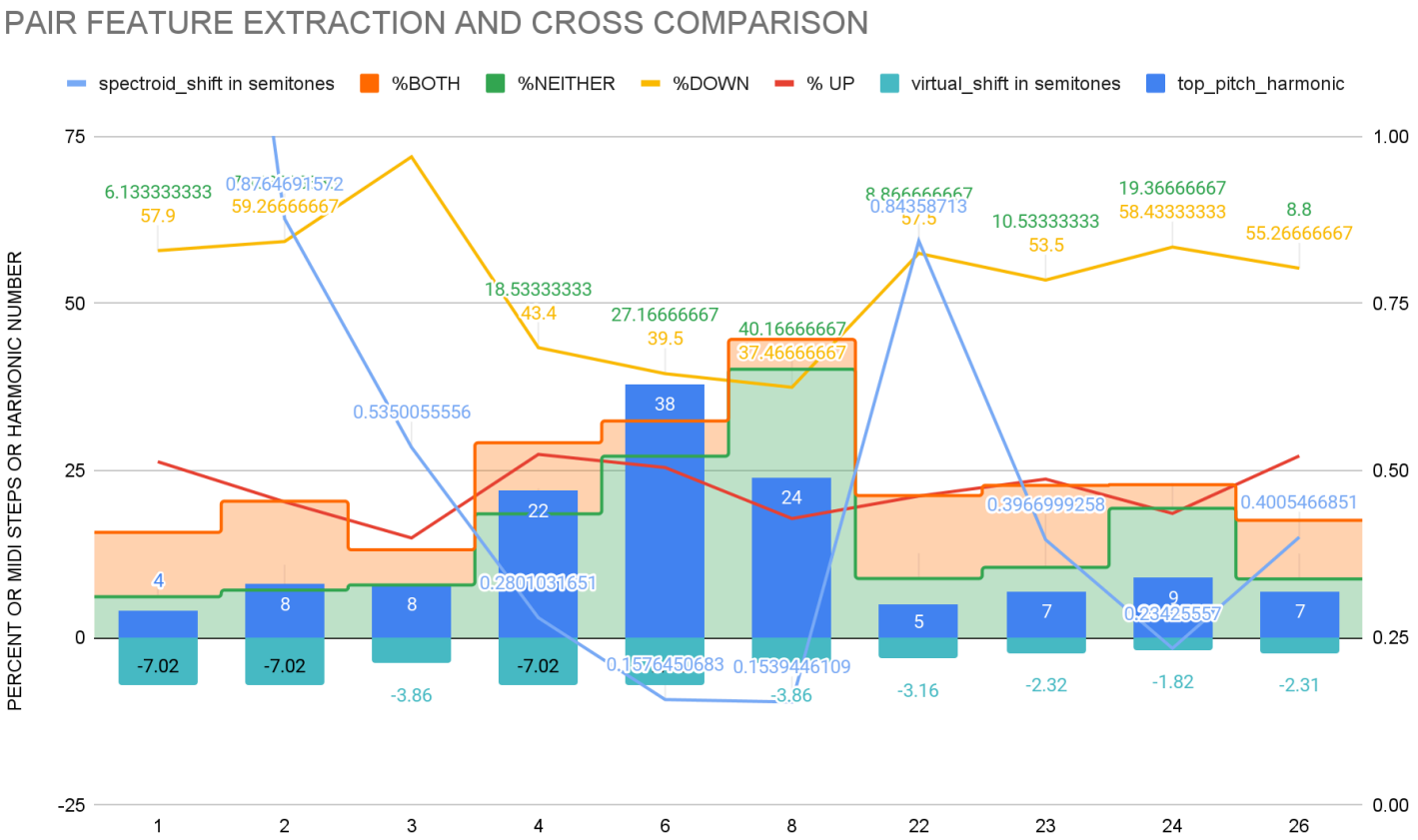
Without starting these analyses, it is worth noting that the stimuli design guaranteed the top frequencies to always be the same. Thus, the minimum inter-harmonic distance is always zero. In other words, it is entirely possible that some responses in the "neither" were describing a listener who was tracking the top partial and hearing no movement, which would not align with the limited paradigm of two mutually exclusive listening modes. Since all tone pairs will show a minimum distance of zero, there is no way to explore results pertaining to this possibility. More informative and complex data collection could be useful for investigating the existence or characteristics of this type of listening.

Figure: Virtual shift in MIDI units (teal) plotted against the left axis. Spectral shift (light blue) in MIDI units plotted against the right axis. Links to sound files can be found in the spreadsheet linked to this figure.

From a visual examination, "neither" percentages seem to be best explained best by shifts that are too small in the spectral centroid. This is presumably for spectral listening only, since the virtual shift is much larger. Another notable trend is the decrease in virtual listening for pairs 4, 6, and 8 along with increases in harmonic rank of partials and rate of "neither" responses.

In fact, the distance of the top partial from the virtual fundamental (blue column) in pairs 4, 6 and 8 might explain the decrease in rate of virtual listening (yellow) for those pairs. One hypothesis is that the distance of the three partials from the fundamental relates to difficulty in

[3] https://github.com/tyfurrier/SpectralF0



listening virtually. That hypothesis could be attributed to the equal spacing being harder to confirm as it is a tiny space relative to much higher frequencies.

If this hypothesis is true, then it could be affecting the neither-rate slightly due to a higher percentage of participants listening spectrally. However, the blue bars are an epiphenomenon. As the partials are higher in harmonic rank, the spectral centroid shift will decrease as the difference in fundamental equates to smaller intervals relative to the higher frequencies. Hence, increasing distance from fundamental to the partials is a cause for a smaller spectral centroid and potentially higher rates of spectral listening.

Visually, the (green) percent of responses in the “neither” category is best accounted for by the spectral centroid shift (blue).

For pitch movement to not be discriminated in the frequency range of the harmonics produced, a pure tone would need to change by around 10 cents or more (Kollmeier, 2008). Further, less movement should be necessary for a complex tone. So, if the spectral centroid is shifting by more than 15 cents in tone pairs 4, 5, 6, and 24, can we surely attribute it to the spectral shift being too small? It could be the case that the visual relationship between spectral centroid and percent of responses marked "neither" is merely a coincidence or a product of confirmation bias, not an increased probability of hearing no pitch movement. Rather, these could be products of the third mode of listening as discussed earlier, when a listener is tracking two individual partials that are the same in each tone.

Distances could be too small because participants are not listening as actively as those in the JND studies. Alternatively hypothesizing, there is an effect of the tone pair only having three partials, complexity, lack of *full spectral smoothness* [4] or its synthetic nature that potentially causes listeners to engage differently and have a higher tolerance for pitch deviation similar to the decrease in JND for speech when compared to pure tones(Johan, 1981). Quick searching into psycho-acoustical studies of pitch discrimination could quickly debunk or deepen that possibility. These are just theories for the moment but overall it raises questions about how listeners are engaging with these tones when compared to typical complex tones.

Limitations / Next Steps

With easily scrutinized steps for ensuring constant volume across remote participants, one might wonder: What would an in-person version look like? Could the hypothesis be retested and show differing results? Based on the peculiarities of the stimuli used and the respective

[4] In other words, these tones do not contain the full continuous harmonic series and the inconsistency could plausibly disconnect the listener from the musicality of the sound object. This could be worth looking into.

responses in “both” and “neither”, a can of worms may have been opened and it seems that this research could have room for further development. With the introduction of the options of “both” and “neither” responses, we found evidence of complications that may have been overlooked by previous studies. Further, these complications could potentially lead to a paradigm shift in the way this research question is fundamentally defined.

We cannot assume the meaning behind “both” responses recorded by participants. Listeners could be switching between virtual and spectral listening for the same pair, they could be tracking individual partials rather than the centroid of the tone, or, this could be simply misunderstanding the prompt or idea of pitch. Responding “both” does not definitively tell us that there is an unanticipated perception going on, that listeners are switching, nor that listeners are simultaneously hearing both spectral and virtually. “Neither” responses on the other hand point directly to a new interaction.

“Neither” had many responses even though the original intent was for it to be an unused option to catch participants who were faking responses or in case something was going wrong. It turned out that things were indeed going wrong. Certain tone pairs had significantly high proportions of “neither” responses. It seemed that some unanticipated effect of confusion or inability to distinguish between the tones was occuring, which could have potentially been understood from an optional text entry for clarification. Based on the variability in selected tone pairs, initial speculation led the research team to investigate hypothesized indicators for high “neither” rates.

The “both” data points were not closely examined, but if the study was replicated with the subjects singing what they heard, this could verify if they are able to hear the pair both virtually and spectrally and count those responses to gain more insight as to when pairs are

perceived through differing mechanisms. This singing approach could help resolve the unwanted percepts that this study obtained, if not discover new models and important questions.

If singing replications and investigations into alternate modes of listening are fulfilled, the parameters for forcing the virtual/spectral dilemma can be reinforced. It could be crucial that any effect explaining the confounding stimuli traits shown here is fully understood before proceeding, rather than just settling on a set of stimuli that does not statistically significantly exhibit symptoms of the effect. This way, models and theories can be developed to better understand the neural foundations behind the listening that this paper is concerned with, leading to more informed exploration, greater connections, and removal of uncertainty that could (has) otherwise go unnoticed. Only once the parameters surrounding this research question have been responsibly revised, would it make sense to retest the loudness hypothesis in more controlled environments.

Further exploration should account for reaction to stimulus, including animosity towards the harshness of tone pairs, which could affect the latency of response time and incorrect selection. There may also be bias introduced by familiarity with the tone pairs. Since different frequency ranges of the same tone pair were used in the stimuli, it is possible this could have primed some participants, especially if they were musically trained.

In a laboratory setting, concerns could be addressed around variability across the listening apparatuses of different users. Some bias could have been introduced in the survey format, with individuals differing in their speakers and hardware, which could result in different frequency responses and hardware clipping. Additionally, no physical tests were performed in a controlled laboratory setting to ensure proper hearing, and normal ear health, including an absence of excess cerumen. In a lab setting, actual pressure readings could be obtained to ensure

a hardware setup with an adequately flat frequency response and no hardware clipping at the required volumes. A more interpersonal and involved procedural environment could see bulletproof steps in ensuring the calibration of hearing thresholds.

References

Belfi, A. M., & Loui, P. (2020). Musical anhedonia and rewards of music listening: current advances and a proposed model. *Annals of the New York Academy of Sciences*, *1464*(1), 99–114. https://doi.org/10.1111/nyas.14241

Chau, Bolton K.H.a,b; Jarvis, Huwc; Law, Chun-Kita; Chong, Trevor T.-J.c. (2018) Dopamine and reward: a view from the prefrontal cortex. Behavioural Pharmacology 29(7):p 569-583, October 2018. | DOI: 10.1097/FBP.0000000000000424

Coffey EBJ, Colagrosso EMG, Lehmann A, Schönwiesner M, Zatorre RJ (2016) Individual Differences in the Frequency-Following Response: Relation to Pitch Perception. PLOS ONE 11(3): e0152374. https://doi.org/10.1371/journal.pone.0152374

Dai, Huanping. (2010). Harmonic pitch: Dependence on resolved partials, spectral edges, and combination tones, Hearing Research, Volume 270, Issues 1–2, 2010, Pages 143-150, ISSN 0378-5955, https://doi.org/10.1016/j.heares.2010.08.002.

Floresco, S.B., Magyar, O. (2006). Mesocortical dopamine modulation of executive functions: beyond working memory. Psychopharmacology 188, 567–585 (2006). https://doi.org/10.1007/s00213-006-0404-5

Furrier, T. (2024). *SPECTRALF0*. GitHub. https://github.com/tyfurrier/SpectralF0

Kiang NYS, Sachs MB, Peake WT. (1967). Shapes of tuning curves for single auditory-nerve fibers. *J Acoust Soc Am.* 1967;42:1341–1342.

Gepshtein, Sergei, Li, Xiaoyan, Snider, Joseph, Plank, Markus, Lee, Dongpyo, Poizner, Howard. (2014). Dopamine Function and the Efficiency of Human Movement. *J Cogn Neurosci* 2014; 26 (3): 645–657. doi: https://doi.org/10.1162/jocn_a_00503

ISO 389-7:2019(en) Acoustics — Reference zero for the calibration of audiometric equipment — Part 7: Reference threshold of hearing under free-field and diffuse-field listening conditions. ISO. (2019). https://www.iso.org/obp/ui/en/#iso:std:iso:389:-7:ed-3:v1:en

ISO. (2023). Acoustics — Normal equal-loudness level contours. Vernier, Geneva. https://cdn.standards.iteh.ai/samples/83117/6afa5bd94e0e4f32812c28c3b0a7b8ac/ISO-226-2023.pdf

Ladd, D. R., Turnbull, R., Browne, C., Caldwell-Harris, C., Ganushchak, L., Swoboda, K., Woodfield, V., & Dediu, D. (2013). Patterns of individual differences in the perception of missing-fundamental tones. Journal of experimental psychology. Human perception and performance, 39(5), 1386–1397. https://doi.org/10.1037/a0031261

Marsh RA, Nataraj K, Gans D, Portfors CV, Wenstrup JJ. (2006). Auditory responses in the cochlear nucleus of awake mustached bats: precursors to spectral integration in the auditory midbrain. J Neurophysiol. 2006;95:88–105.

Meddis, R., & Hewitt, M. J. (1991). Virtual pitch and phase sensitivity of a computer model of the auditory periphery: II. Phase sensitivity. Journal of the Acoustical Society of America, 89(6), 2883–2894. https://doi.org/10.1121/1.400726

Micheyl, C., Delhommeau, K., Perrot, X., & Oxenham, A. J. (2006). Influence of musical and psychoacoustical training on pitch discrimination. Hearing research, 219(1-2), 36–47. https://doi.org/10.1016/j.heares.2006.05.004

Micheyl, Christophe, Oxenham, Andrew J. (2010). Pitch, harmonicity and concurrent sound segregation: Psychoacoustical and neurophysiological findings, Hearing Research, Volume 266, Issues 1–2, 2010, Pages 36-51, ISSN 0378-5955, https://doi.org/10.1016/j.heares.2009.09.012.

Nieoullon, Andréa; Coquerel, Antoineb. (2003). Dopamine: a key regulator to adapt action, emotion, motivation and cognition. Current Opinion in Neurology 16():p S3-S9, December 2003.

Oxenham, AJ. (2012).Pitch perception. J Neurosci. 2012 Sep 26;32(39):13335-8. doi: 10.1523/JNEUROSCI.3815-12.2012. PMID: 23015422; PMCID: PMC3481156.

Schneider, A. (2018). Pitch and Pitch Perception. In: Bader, R. (eds) Springer Handbook of Systematic Musicology. Springer Handbooks. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-662-55004-5_31

Schneider, P., Scherg, M., Dosch, H. G., Specht, H. J., Gutschalk, A., & Rupp, A. (2002). Morphology of Heschl's gyrus reflects enhanced activation in the auditory cortex of musicians. Nature neuroscience, 5(7), 688–694. https://doi.org/10.1038/nn871

Seither-Preisler, A., Johnson, L., Krumbholz, K., Nobbe, A., Patterson, R, Seither, S., & Lütkenhöner, B. (2007). Tone sequences with conflicting fundamental pitch and timbre changes are heard differently by musicians and nonmusicians. Journal of Experimental Psychology: Human Perception and Performance, 33, 743–751. https://doi:10.1037/0096-1523.33.3.743

Steinschneider, M. (2011). Unlocking the role of the superior temporal gyrus for speech sound categorization. *Journal of Neurophysiology*, *105*(6), 2631–2633. https://doi.org/10.1152/jn.00238.2011

Winer, Jeffery A. (1984). The human medial geniculate body, Hearing Research, Volume 15, Issue 3, 1984, Pages 225-247, ISSN 0378-5955, https://doi.org/10.1016/0378-5955(84)90031-5.

Yuskaitis CJ, Parviz M, Loui P, Wan CY, Pearl PL. (2015). Neural Mechanisms Underlying Musical Pitch Perception and Clinical Applications Including Developmental Dyslexia. Curr Neurol Neurosci Rep. 2015 Aug;15(8):51. doi: 10.1007/s11910-015-0574-9. PMID: 26092314; PMCID: PMC5469678.

Zatorre R. J. (2005). Neuroscience: finding the missing fundamental. Nature, 436(7054), 1093–1094. https://doi.org/10.1038/4361093a

Zatorre, R. J., & Belin, P. (2001). Spectral and temporal processing in human auditory cortex. Cerebral cortex (New York, N.Y. : 1991), 11(10), 946–953. https://doi.org/10.1093/cercor/11.10.946

't Hart, Johan (1981). "Differential sensitivity to pitch distance, particularly in speech". *The Journal of the Acoustical Society of America*. **69** (3): 811–821. Bibcode:1981ASAJ...69..811T. doi:10.1121/1.385592. ISSN 0001-4966. PMID 7240562.

Kollmeier, B.; Brand, T.; Meyer, B. (2008). "Perception of Speech and Sound". In Jacob Benesty; M. Mohan Sondhi; Yiteng Huang (eds.). *Springer handbook of speech processing*. Springer. ISBN 978-3-540-49125-5.