Effect of Infra-Harmonic Pitch Distribution on Timbral Perception and Evidence of abstract coding for memory

Tyler Furrier

Department of Music, Northeastern University

MUSC3300: Music Perception and Cognition

Dr. Psyche Loui

December 12, 2023

Abstract

In this study, we explore the question of spectral resolution with concern to timbral distinction. We asked participants to listen to two sounds of the same timbre differing only in the pitch width of each partial. Current research has shown augmented results on pitch discrimination and scrutinized the notion of critical bands when considered in the context of timbre perception, especially harmonic sounds. As expected, we found that the threshold for discrimination is smaller than that found in typical studies of critical bandwidth and pitch discrimination. Our results may suggest that if the critical bandwidth still applies, there is a perception of the bandwidth widening. The results suggest that the spread out pitches were cognitively grouped to their center, along with some form of memory of the *scale* of their spread in a timbral characteristic. Notably, distinctions were best predicted by the ratio of widths between the two sounds. This potentially indicates that we dynamically adjust our resolution of pitch perception, like zooming a camera lens, depending on the sonic environment we are observing. If that is the case, this study serves as evidence that the accuracy of pitch perception is relative to the pitch range being compared.

Introduction

The study of timbral perception is one spread wide and far. Like many questions of psychology and neuroscience, studies of various features and their various effects tease at a centralized theory that explains our perception of timbre. A goal like this is becoming all the more pertinent with rises in technological opportunities in MIR and musical human computer interaction. However, sometimes scientific pursuits can be lead away from answers when we take for granted or generalize previous findings. Most theories of timbre revolve around a

sound's complexity, harmonicity, temporal envelope, and spectral distribution on a macroscopic level. In this paper, we seek to better understand the spectral resolution at which we judge a sound's timbre, and at which we segregate timbres. With modern tools like machine learning and generative models, it might be better to focus on defining the microscopic features of timbre, enabling tools to better examine all factors involved in various perceptual, psychological, physical, or even subjective outcomes.

I could throw a stone and it would land on a research paper considering the macro spectrum of a sound, but I struggled to find research considering the content in and around each major partial of a sound, ie. the spectral resolution of our musical perception. This could be attributed to foundational psychophysical research around pitch discrimination and the critical band(#4), which would have us believe that pitches within a small enough range will be perceived as the same and we are left only with the measurement of magnitude. I find it highly unlikely that a mechanism as astounding as our hearing and musical cognition would perceive two neighboring notes the same exact way as one at the same combined magnitude. Further, there is evidence that performance in pitch discrimination tasks changes for stimuli of complex tones and real sound objects, indicating different thresholds than those inferred by single tone studies(#1).

I'm not going to rant as if my naive view of music perception and cognition outweighs the years of research. Critical bandwidth does obviously exist, and within a small enough interval we can't differentiate pitch. We can only perceive Xct difference in pitch. X might change depending on frequency or various factors of the sonic environment around it, but there still is a reasonably large X. Does that imply we can only perceive Xct pitch bins of magnitude when

perceiving a timbre? Rendering our perception of each partial as no more than a fancy curvy rectangle with a width of X?

For starters, the box is not fixed(#5). McAdams studied frequency modulation incoherence, where ratios among the component frequencies are not constant. His findings showed that for harmonic sounds, the auditory system can detect incoherence at modulation widths of less than 0.05%, much less than the ~20% critical bandwidth(#4). With this, it becomes clear that we are physically capable of detecting smaller differences. It is then a question of when we are cognitively able to perceive them. Answers considering modulation still leave questions as to whether these differences are perceived in steady-state signals, when the different sounds are not played back to back. McAdams and colleagues answer this question as well(#3). By tasking subjects with matching the pitch of a mistuned harmonic, they found that general ability to perceive inharmonicity "augments" the data obtained from discrimination experiments and argues against models based solely on partial masking. They also found a decrease in ability in higher frequency ranges "where synchronous neural firing vanishes". These studies from McAdams open the floor for novel questions, but still only consider movement of partial pitches. In other words, if the mistuned harmonic is creating a critical bandwidth, the center of the bandwidth is detuned. In this study, we examine the widening of a complex harmonic tone's partials by combining various neighboring frequencies to create the same perceived magnitude and pitch center but with a wider composition.

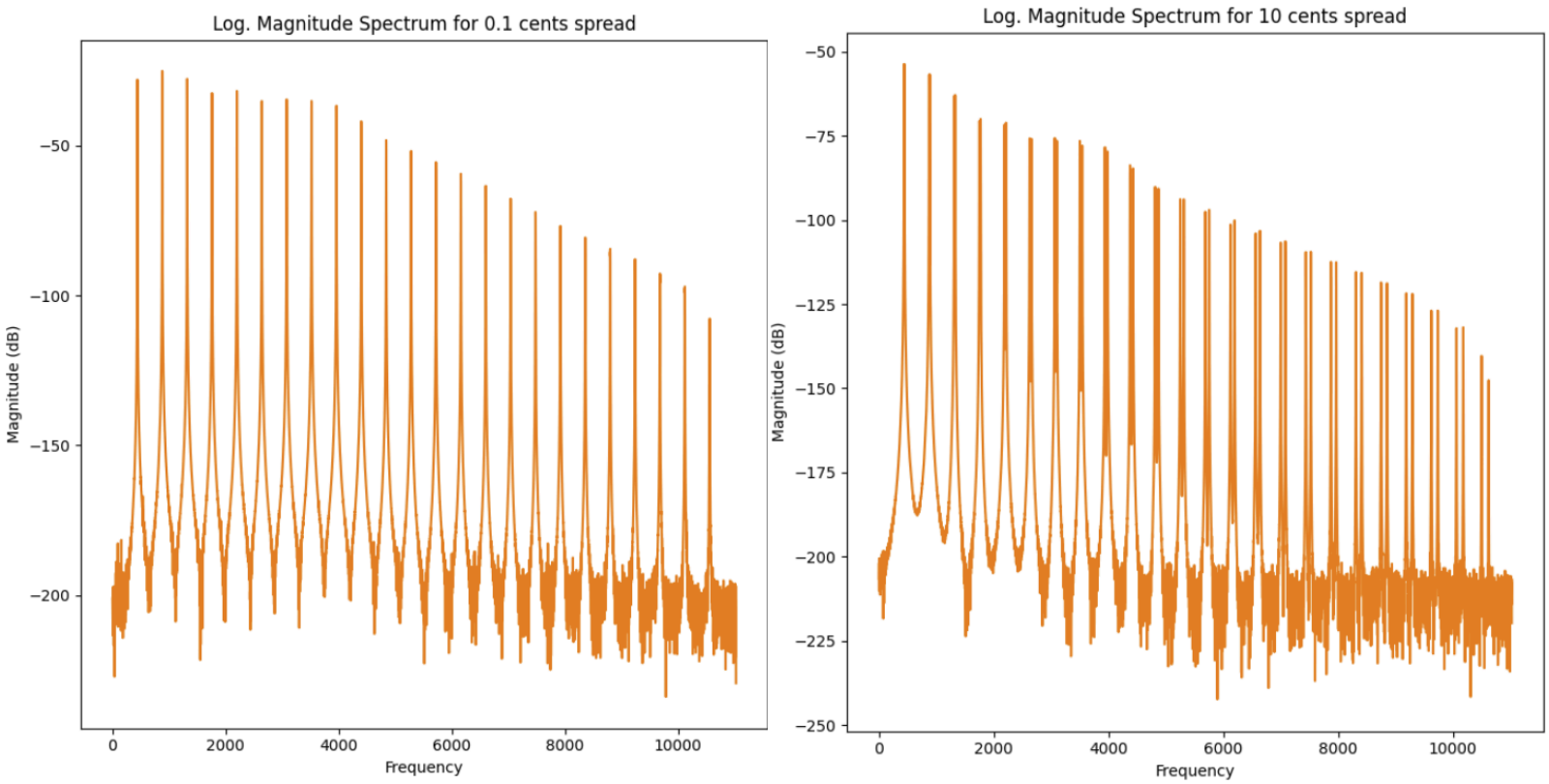
Methods

Subjects

Results were contributed by a mix of 6 coworkers. Their musical backgrounds or skills are unknown, but one subject had extremely erratic data and his responses were thrown out. With

the study starting small, we planned inclusion of participants with variance in musical skill aims to capture a holistic perspective on the generality of our data and how to best recruit a second batch of subjects. In reality, we had no say in the musical significance of each participant, but one could make the argument that we better represented a random sample of the entire population by avoiding musically influenced circles.

Figure 1: Spectrograms of different stimuli



Stimuli

We aim to present varying pairs of nearly identical sounds made through additive synthesis. The only difference between these sounds was the local width in pitch of each harmonic or partial.

In practice, realizing this vision was quite difficult and is still not perfect. The first generation of sounds created a "radius" around the harmonics by combining pure tone synthesization of the sound whose fundamental frequency falls within the radius of the original fundamental frequency. An N cents radius synthesization for f_0 was created by summing the pure-tone synthesizations[1] from f_0 - N to f_0 + N where each component was divided by 2N + 1. Notably, the level of granularity was only one cent. Beating, along with a couple other sneaky quirks[2], was creating artifacts so severe that the task became one of rhythm memorization or masking quick salient sounds. I could not use any data from thanksgiving. After many changes in approach, moments of self-doubt, and batches of wav files, I landed on a proper set of sounds that isolated the percept well. The sounds were generated the same way but with almost (mentioned in considerations) no artifacts thanks to more precise pure-tone generations and

[1] An additive synthesis attempt at a trumpet but only using raw pure tones for each harmonic

[2] An optimization declaring the amplitude of a harmonic as the current bit index times the harmonic % signal length meant crackles and pops were occurring each time the index wrapped back to the front. This presumably occurs because the one second wave being generated was not a perfect loop. Harmonics were shooting up and down at different times in order causing a wild effect. To add insult to injury, I kept the index zero-indexed meaning quality was presumably also deteriorated as each fn amplitude for a given bit was off by n bits. Hear it yourself in the created_samples/unused-old/440_spread_full folder

intervals of 0.1 cents compared to the 1 cent interval used prior. The dB spectrum for some of these sounds can be seen in figure 1.

The sounds still had an initial dropoff when phase cancellation occurs most aggressively, creating a shorter attack time for sounds with higher radii. Further, wider radii resulted in lower volumes. For our next tests, we will generate the sounds for longer and crop off starting "pings" and then normalize the volume of the resulting sounds. For pilot results, we instructed participants to ignore differences in volume and attack, focusing only on the texture of the sustain. They seemingly did a good job at that and did not cheat. However, if results change after kneading out the knots, we will be sure to replace their data.

The subjects completed steps in which they were presented two sounds, each with a random radius from 0 to 499 *tenths* of a cent. While staircase methods do not select comparisons randomly, we opted for random to increase the number of data points since it meant zero time commitment. Further, the results indicate the perceptible difference is best described as a proportional difference in cents rather than absolute difference. Thus, it is imperative that various floors and ceilings for radius be used. Consequently, the linearly random distribution does not distribute evenly across the perceptual distance between sounds. The next iteration of tests will attempt to weigh the random selections logarithmically. The subjects had the choice of hearing the sounds as many times as they wanted.

Finally, for each comparison, we have also generated sounds for a comparison where only some partials are widened[3]. The partials that are widened in this case were those present in the harmonic series of a note one octave higher than the original pitch.

Procedures

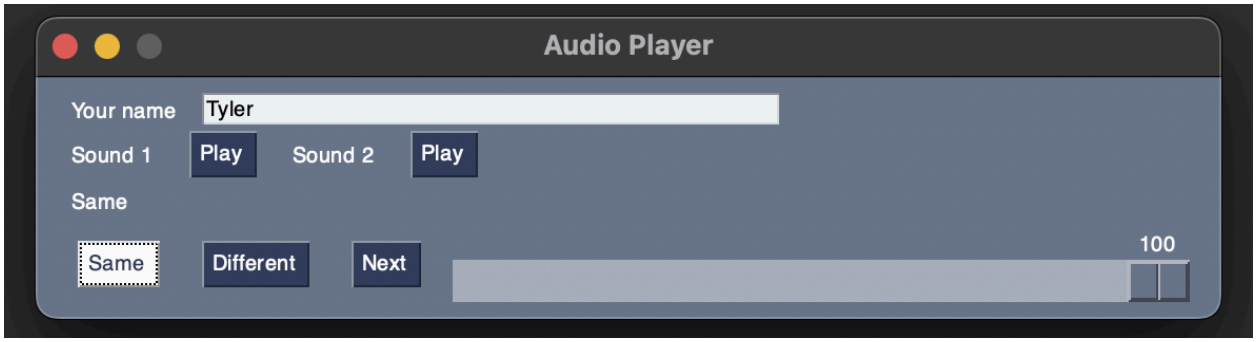
[3]By the time I needed results, volume differences between radii were too severe for an ideal sharing of harmonics. Also, I wanted to prioritize data for experiment 1 first.

Before starting the experiment, a pre-survey gauged participants' current state and recent music or sound exposure[4]. Then participants heard a 0.1 cent radius sound, 2 cent radius sound, and 20 cent radius sound so they had a context in which to judge similarity of the sounds. The participants were also told to compare only the textures of the sounds and ignore anything else like volume and temporal effects like a perceived attack caused by the hasty sound generation. This initial introduction listening, paired with the survey, aims to control for any unforeseen confounding factors that different timbral or musical contexts cause.

For all comparisons, the subjects were asked whether the timbres are the same. To better quantify cases close to the perceptible difference with a small number of participants, the amount of times a sound was played and the time to answer was tracked. In future comparisons with combinations of an octave and original at different radii, there will be an additional prompt to select if either of the sounds have more than one "instrument" present. In further testing, to raise the likelihood of the data being reusable (or flexible if we find unexpected results or an answer to a more appealing hypothesis) the subjects will also have the opportunity to provide optional responses of the difference in various timbral characteristics between the two sounds. For accurate difficulty tracking, these extra prompts will be hidden away on a page after the main page.

Figure 2: The GUI used for the experiment

[4] For pilot results, all had been in the same office listening to the same music. Conveniently, they were working all day so this may have been their first active listening in a while which hopefully meant their ears were fresh!



Participants all used the same headphones(audio technica ATH m50x) plugged into a Komplete audio 6 interface which was connected to my laptop. Subjects interacted with a graphical user interface (GUI), shown in figure 2, that tracks all of the user's actions, synchronized with their computer audio, microphone, and potentially camera. Volume was fixed until the second subject mentioned he could not hear anything after the very beginning of a 49ct radius sound, at which point the volume was increased. The GUI allowed the sounds to be played quickly by either a mouse press or keyboard press of the 1 or 2 key. Answer changes were allowed, but once the "next" button was hit the answer was locked. By forcing the user to finalize their same/different answer before answering more, we have allowed ourselves to more precisely track the number of answer switches, sound play times, and overall time to answer the question. Any behavioral analysis on this data like time-to-answer will be looked at on both a within-subjects basis and between-subjects basis appropriately. If the study is to be replicated on a larger scale, the ordering of these prompts and the likelihood of seeing optional prompts could be randomized.

After the experiment, subjects made a self-assessment and reflection to gauge where they were listening on a range from most actively to most passively. They are all nice people who care about my results and I gave them an imaginary sense of competition by making judgements after

they were done. Hence, they all answered that they listened very carefully and had various ranges of confidence.

Results

Figure 3: scatterplot of responses based on difference in cents/10 width of two sounds, ratio of widths, and log ratio of widths

Data Analysis

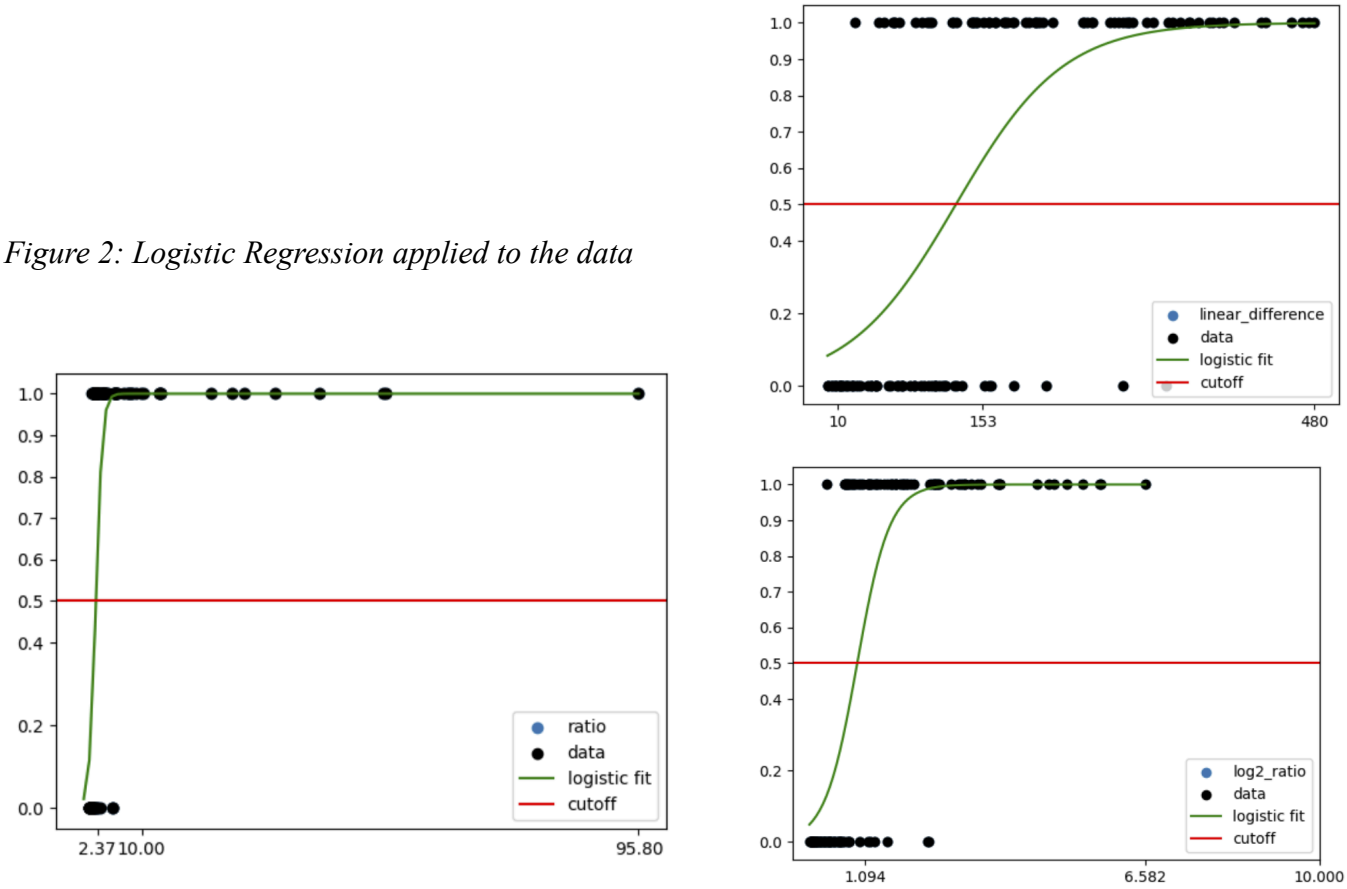
The experiment being conducted through a Python-based app, we collected behavioral and time-based data from each session. The results contain rows for each comparison in a csv which can be seen on github at:

https://github.com/tyfurrier/MPC-research/blob/main/spreadsheet.csv

It will be useful to conduct d prime tests in various ratio bins, but our current batch of data is too spread out to create well sampled clusters of nearby results for a given ratio. In lieu of a d prime visualization, we fit a cumulative Gaussian distribution via SciKit-Learn's LogisticRegression object in Python.

After fitting models to the ratio, log ratio, and linear difference between widths, all three scored highly with mean accuracies of 0.853, .838, and .823 respectively. Logistic regression showed a 0.02 p value for ratio and 0.07 for linear difference.

Figure 2: Logistic Regression applied to the data



We still have a potential question to answer. Is there a pitch size small enough in width that we are unable to distinguish it from any smaller width with the same cumulative amplitude? Shifting our attention to the linear scale of figure 1, our data suggests that any difference smaller than 4 cents (4 cents diameter) is indistinguishable. However, absolute claims like a fine cutoff can not be made until this study is replicated on a larger scale and with better stimuli.

Discussion

If differences are perceived, then we ought not to over generalize the concept of an equivalent rectangular bandwidth as causing neighboring frequencies to not be perceived, especially in the discussion of timbre. While subjects didn't categorize the sounds as being different pitches, it is evident that the brain encoded a perceived widening of the excited area across the membrane. It's safe to say, for comparisons with a low ratio in widths of partials, subjects did not remember the sounds as having a difference in pitch width or dissonance. Rather, the subjects, as directed, has *memory* of the texture. Similar to efficient encoding presented by Malinda McPherson(#2), this abstract characterization of different distributions of psychophysical frequency stimulation seems indicative of a relationship (between our variable and timbre) and an abstract deduction of texture by which the stimuli differences were more efficiently memorized on a logarithmic scale.

These results do not contradict the physical idea of critical bands. One way to see if critical bands are not aptly suited for representing our perception would be to create sounds that result in the same band and check for distinction. I'm not currently sure if that's possible

Visually looking at the data, there seems to be a blurry area where a cutoff is not as distinct as we hoped. With that considered, next steps include gathering more data, computing data within subjects, and then using more involved ML methods to understand if the detection cutoff is better described in context of the widths compared rather than just the ratio in width between the two. For example, the difference between a 1 and 2 cent spread has a clear timbral change. However, 0.1 and 0.2 cents have the same ratio but sound identical. Similarly, 25 cents and 50 cents are wide enough that they sound similar. Applying various algorithms may help find a formula that defines borders of perception and the residual effect around them if it is present.

https://github.com/tyfurrier/MPC-research

References

1

McPherson, M. J., & McDermott, J. H. (2018). Diversity in pitch perception revealed by task dependence. *Nature human behaviour*, *2*(1), 52–66. https://doi.org/10.1038/s41562-017-0261-8

2

McPherson, M. J., & McDermott, J. H. (2020). Time-dependent discrimination advantages for harmonic sounds suggest efficient coding for memory. *Proceedings of the National Academy of Sciences*, *117*(50), 32169-32180.

3

William Morris Hartmann, Stephen McAdams, Bennett K. Smith; Hearing a mistuned harmonic in an otherwise periodic complex tone. *J. Acoust. Soc. Am.* 1 October 1990; 88 (4): 1712–1724. https://doi.org/10.1121/1.400246

4

Gelfand, S. A. (2004). *Hearing: an introduction to psychological and physiological acoustics* (4th ed.). New York: Marcel Dekker. ISBN 978-0-585-26606-0.

5

Mcadams, Stephen. (1984). Spectral fusion, spectral parsing and the formation of auditory images. [Doctoral dissertation, Stanford]