Running Head: ANALOGIES FROM SPEECH TO MUSIC

THE UNIVERSITY OF CHICAGO

Analogies from Speech to Music: Processing Costs of Switching Between Timbres in Musicians Without Absolute Pitch

By Matthew C. Di Santo

August 2023

A paper submitted in partial fulfillment of the requirements for the Master of Arts degree in the
Master of Arts Program in the Social Sciences

Faculty Advisor: Howard Nusbaum

Preceptor: Tori Gross

Abstract

Prior research has shown analogies between speech and music, and musicians who possess absolute pitch have been studied extensively. To generalize these findings, active musicians who did not possess absolute pitch completed a speeded pitch recognition task in which a target pitch was identified among trials blocked or mixed instrument timbres. Hit rate and reaction time were measured and compared by trial type. It was hypothesized that mixed trials would elicit lower hit rate accuracy and slower response times than blocked trials. There was no significant difference between the mean hit rates or response times for blocked trials vs. mixed trials. Future research should increase the sample size of the study and include non-musicians for greater generalizability of results.

Analogies from Speech to Music:
Processing Costs of Switching Between Timbres in Musicians Without Absolute Pitch

Many aspects of our auditory processing depend on the context that surrounds particular stimuli. Within the realm of auditory stimuli, there are two groups that share many analogous features: speech and music. Speech and music, or more specifically speech and instrumental music, possess a variety of acoustic elements that we can use to distinguish between different talkers or different instruments. There are two aspects that are most important to the current study. The first is a particular talker's quality of voice and vowel space which is analogous to a particular instrument's timbre. Quality of voice and vowel space refer to the particular sound of a person's voice as well as the production of various formants within the particular spaces in the oral cavity that create vowels (Nusbaum & Morin, 1992). Timbre refers to the particular and unique sound of one instrument from another that we can use to differentiate between them. The second aspect of importance to this study is the fundamental frequency and range of one's voice, being analogous to the octave range of an instrument. The fundamental frequency of one's voice refers to the pitch at which a talker most normally speaks, and the range that they exist within either direction of this pitch (Magnuson & Nusbaum, 2007). The octave range of an instrument on the other hand, refers to the range of notes that an instrument is able to produce measured by octaves, or the scalar eight-note repeating sequence used in Western classical music.

Speech and Different Talkers

As previously described, there exist many analogies between the auditory stimuli of speech and instrumental music. One of these similarities relates to acoustic features that are similar between a speaker or individual instrument and how we subsequently process them. Nusbaum and Morin (1992) conducted a series of experiments to determine the various ways in which we form a normalization model for different talkers. One important experiment consisted of a speeded recognition task in which 16 stimuli were separated by 250 ms with a target vowel to be identified among a group of distractor vowels. Participants had to press a button as quickly as possible to indicate whether or not the vowel heard was the target. This was presented in blocked and mixed trials where blocked used the same talker and mixed used different talkers. Reaction times for vowel recognition were slowed in mixed trials when compared to blocked. They additionally found that within the same talker, we use a variety of cues within a set of speech to predict and form a prototype to normalize a talker's voice, and that this normalization process draws heavily on different contextual elements of a talker's speech; a process the authors name "contextual tuning" (Nusbaum & Morin, 1992).

This normalization model (Nusbaum & Morin, 1992) can help us further understand how we may then cognitively process the speech of different talkers. In a later experiment also by Nusbaum, Magnuson and Nusbaum (2007) conducted a series of experiments examining processing costs related to the perception of different talkers' speech. The main experiment of interest found that in a speeded recognition task similar to Nusbaum and Morin (1992), there was a main effect of trial type found in which hit rates for target vowels in blocked conditions were higher than for mixed conditions. Additionally, response times for mixed conditions were slower than for blocked. In the real world, this translates to the fact that the processing and recognition of a target vowel measured by response time (seconds) was significantly slower when two talkers differed significantly in their vowel spaces and/or their fundamental frequencies. This finding relates to the aforementioned study (Nusbaum & Morin, 1992) related to contextual tuning as it shows that if we form a normalization for a particular talker, this can produce temporal processing costs when we later want to switch between talkers.

Analogies from Speech to Music

Many studies that hope to reveal the underlying processes of music cognition test the ability of a listener with absolute pitch to accurately identify a given note in a Western classical scale across a variety of musical stimuli. Absolute pitch refers to an individual's ability to identify or produce one of the 12 notes present in Western classical music without the need for a reference pitch or other context. Bahr et al. (2005) conducted a study in which participants with absolute pitch who played a variety of instruments had to correctly name a played pitch. They found that the timbre of the instrument that one had significant experience playing as well as the range of pitches it could produce had a significant effect on a participant's pitch identifying accuracy (Bahr et al., 2005). This relationship existed such that a participant was better able to correctly identify a played pitch when the timbre and octave range of the instrument it was played in closely aligned to the instrument they had experience playing.

Still regarding instrument timbre, Brammer (1951) found that violinists experienced more difficulty tuning a played note to a target pitch measured in cents when it was played on a clarinet rather than a violin. Cents refers to the ability of a given pitch to be divided into one hundred subintervals. This means that when violinists tuned the played note to the target pitch, they were further away in musical cents from the accurate pitch when tuning a clarinet as opposed to a violin. These experiments show that both instrument timbre and octave range in music, being analogous to voice quality, vowel space, fundamental frequency and range in speech, have significant effects on one's ability to process these auditory stimuli with differences between certain variables incurring processing costs.

Processing Costs in the AP Population

Van Hedger et al. (2015) used the previously presented analogies between speech and music to inspire their study in which absolute pitch possessors had to correctly identify a target pitch among a variety of contexts. The researchers referred to Nusbaum and Morin's (1992) study and the previously described contextual tuning theory to draw and support the use of such analogies between speech and music. In their study, Van Hedger et al. (2015) conducted a series of experiments, with one of note in particular. In this experiment, a series of notes were presented to the absolute pitch-possessing individuals. These notes were 500 ms in duration, and were presented with 250 ms of silence between each note. One target note was used that was placed in a set of 16 total stimuli per condition, with three other notes serving as distractors (4 targets total, 12 distractors), and the participants had to identify using a button press when they heard the target note in each trial. The trial had a pseudo-random presentation, meaning that target notes were never presented in the first or last position and were also never presented back-to-back. Participants listened to two blocked trials and two mixed, in this case, one piano blocked, one violin blocked, and two piano/violin mixed. The researchers found that there was a significant effect of trial type in this experiment, with mixed trials having significantly higher response times than blocked trials (Van Hedger et al., 2015). Similar to Magnuson and Nusbaum's (2007) findings with talkers, this significantly higher response time when switching between instrument timbres and identifying the target tone represents a processing cost as measured by slower reaction times (Van Hedger et al., 2015).

The Current Study and The Importance of the non-AP Musician Population

The current study hopes to replicate the Magnuson and Nusbaum (2007) and Van Hedger et al. (2015) study using a speeded-target recognition task with multiple instrument timbres presented across a range of sine-wave tones in which participants will need to identify a target tone and see if a processing cost is incurred when the timbre changes. Also different from Van Hedger et al. 2015, participants in this study will be musicians without absolute pitch in hopes that the results can generalize to a larger population rather than a select number of musicians who also happen to possess the ability of absolute pitch. I hypothesize that even within musicians without absolute pitch, the identification of target tones will incur a processing cost represented by a longer response time when a timbre change is present.

This experiment is necessary since in the field of music cognition to date, the absolute pitch population has been widely used for a variety of experiments. However, this group represents a very discrete and hardly robust sample of the population of people that are musically trained. There has also more recently been much debate over the ability to characterize the trait of absolute pitch possession as solely dichotomous. For example, Van Hedger, Villeitte, Heald and Nusbaum (2020) tested this exact notion in a study in which 152 participants were asked to identify the name of note when presented in a continuous block of 48 stimuli that spanned two octaves. Between each trial the octave range changed to mitigate the use of any relative pitch strategies. Relative pitch is a colloquial term used to describe musicians who possess the ability to identify one note by the interval distance when it is presented in reference to another note. In this study, the results of the participants were divided into three groups; those who possessed genuine AP, those who did not, and an in between category that the researchers labeled as "pseudo-AP possessors" (Van Hedger et al., 2020). This in between category was made for those who correctly identified at least 11 of the 48 notes by name, and those who identified 39 or above were classified as genuine AP while those who correctly identified 10 or under were non-AP. Importantly, there was a significant difference between the age of musical instruction onset between both the genuine and pseudo-AP groups when contrasted to the non-AP group, in that the age of musical instruction was much earlier for the first two groups.

This means that it is safe to assume that those in the pseudo-AP group have extensive musical training as they did differ in age of onset from non-AP possessors but did not from genuine AP-possessors. Additionally, it supports the idea that there is a critical period for the development of absolute pitch in individuals, but more importantly, that we may be able to view AP on a spectrum rather than as a discrete variable (Van Hedger et al., 2020). This highlights the importance then of studying highly-trained musicians who do not possess genuine-AP, as it helps further generalize the findings of the current study to a larger population of trained musicians who may possess AP like abilities but have not since been included widely in the music cognition literature as a population of interest or study.

While these results suggest the existence of measuring AP on a spectrum, it should also be noted that Van Hedger et al. (2020) did still use labels such as genuine and pseudo-AP participants to break down the fact that there still does exist a difference in those with musical training who genuinely possess AP and those who show abilities similar to it. This notion is supported by the work of Van Hedger, Heald and Nusbaum (2016). In a study in which participants who either had or did not have absolute pitch (thus treating this as a discrete an dichotomous variable), participants heard two presentations of a line of audio spoken by either a comedian or a robotic voice in which the frequency used to bleep out profane language was identical to its real world presentation or varied by either a semitone or two semitones Van Hedger et al., 2016). In each trial one presentation of the audio dialogue line was the correct frequency of 1,000 Hz, and the second presentation was either one or two semitones sharp or flat. The participants had to identify in each trial which presentation was the correct version of 1,000 Hz. A semitone refers to the smallest interval that can exist between two notes in Western classical music. Results showed that even though non-AP participants did well, AP participants still did significantly better in this task. Due to the size of the population used in this experiment, it is presumable that some of the non-AP participants still would have possessed musical training to a degree, meaning that there was still a significant performance difference between participants who do and do not possess AP but are still trained musicians.

In a second experiment to further interrogate the differences between how AP vs. non-AP participants might be processing the stimuli, Van Hedger et al. (2016) lowered the interval size between the 1,000 Hz tone to 42 cents sharp or flat, meaning under a semitone and thus smaller than any division of pitch used in Western classical music. This was done to see if AP possessors were using the fact that they can associate a pitch with a specific note name to help differentiate if the stimuli they heard in Experiment 1 were correct or not rather than non-AP people who would not be able to process the stimuli in this way. In this second experiment, AP possessors still outperformed non-AP individuals at a significant level, again highlighting that there may be a true difference in the way in which pitch is processed in the brains of an AP individual vs. a non-AP individual, regardless of musical training, which provides all the more reason to pursue the current experiment.

To reduce any ambiguity in the assumption that those within the non-AP group may all lack any musical training, Van Hedger et al. (2016) also completed another experiment using the same stimuli as Experiment 2, but this time with individuals who had musical training and also may or may not have possessed AP. In the task, those with musical training, hereafter referred to as "Experts" (Van Hedger et al., 2016), performed between those with AP and those without it when compared to the previous Experiment 2. Furthermore in a third experiment, Van Hedger et al. (2016) found that based on all of the data from these experiments, those who do possess AP do seem to process pitch differently from the non-AP population, which again, can still include other musicians. All of this information lends more of a necessity for the current experiment to be conducted, as it will help show how a larger population, being that of musicians without AP, may process pitch differently or similarly to AP possessors given that there is reason to believe that AP possessors may simply have superior processing for pitch.

Showing the existence of this processing cost evident in AP people when having to identify the same note between two different timbres, but instead within musicians who do not possess absolute pitch, would further clarify how we process musical stimuli as a more generalized rather than specialized population. If the same effect is found, it would show that these processing costs across timbres do not just occur for absolute pitch possessors who use contextual tuning to create normalizations across instrument timbres for pitch (Van Hedger et al., 2015). Rather, it would show that there are other categorical and perceptual processes at work in the brains of all people that are not necessarily pitch dependent for the ways in which we form normalizations across timbres. This would strengthen the connection between how we process musical stimuli and speech stimuli and further extend the results from the Nusbaum and Morin (1992) and Magnuson and Nusbaum (2007) experiments in the broader realm of auditory stimuli processing. Based on the discussion of the previous literature, I hypothesize that non-AP musicians will have a higher accuracy of identifying the target pitch in blocked rather than mixed timbre trials, and that they will exhibit longer response times when having to choose the target note in mixed rather than blocked timbre trials, thus representing a processing cost of switching between timbres.

Methods

Participants

7 participants total (4 male, 3 female) ages 25-67 were included in the analysis for this experiment. Six participants were not included in the final analysis as two did not fully complete the experiment and four completed the study but their data did not save. This left the previously mentioned 7 participants in the final analysis. Participants were selected through a mass text message sent out to colleagues of the experimenter that were classically trained musicians. Participants in this study played a range of instruments including the piano, violin, viola, cello, guitar, electric bass, clarinet, and voice. The mean level of training in years for all participants was 27.14 years. All participants were asked in advance if they had perfect pitch, which would exclude their participation in the experiment, and no participants possessed this trait. Participants were paid $10 for their time and were given the money either in person or through the mail.

Apparatus/Materials

The experiment was coded in MATLAB (v. 9.13.0 R2022b) and jsPsych (de Leeuw, 2015), and hosted on Pavlovia (see references) using the speeded talker recognition task paradigm from Nusbaum and Morin (1992) and Magnuson and Nusbaum (2007). The instrumental stimuli consisted of a xylophone, harpsichord, plucked violin, harp and piano, which was a mixture between stimuli used in Van Hedger et al. (2015) as well as some extra stimuli made for this experiment by Dr. Stephen Van Hedger. These stimuli were all synthesized .wav files, had strong attacks, were matched in terms of root frequency, length (500 millisecond duration) and octave range (C4-B4), and were amplitude normalized. As this study was done remotely, participants used their own personal computers to complete the experiment.

Procedure

Participants opened the experiment and were brought to a screen welcoming them to the study as well as showing them an informed consent form. They were then asked if they agreed to take the experiment. Further questions asked were whether or not the participant was in a quiet place and if they had removed all distractions. They were instructed to answer these questions honestly as this would affect their experimental performance. Participants were then asked their gender, age, which instruments they played, for how long, and if they were still actively playing. These questions were asked in order to make sure that each participant was qualified to take the study as their answers would determine whether or not they could continue.

Following the introductory survey, participants were given the instructions for the experiment at hand. They were told that they would be played a target note, after which they would hear a series of notes in quick succession, and that their task was to press the key “T” every time they heard the target note in question as quickly as they could. After a training trial, the experiment began.

The experiment itself consisted of 10 total blocks, with 5 including just one instrumental timbre and 5 including a mixture of two instrumental timbres. Within each block there were 10 trials, each one with a different target note to identify. Within each trial, participants were presented with a total of 16 notes; 4 of the notes heard were the target and 12 were distractors.

The distractors were chosen to be one semitone apart from the target in either direction. This was to ensure that the experiment itself was difficult enough for trained musicians to complete and to avoid a ceiling effect, as if the distractors were too far apart from the target the task would be too easy and the target would be obvious. The presentation order of the blocks was randomized so that there was no effect of presentation bias in the experiment, and there were different experimental presentations of the stimuli to ensure that not every participant heard the same combination of blocked or mixed instruments. Within each trial, the participant would hear the target note and then press a key when they were ready to continue, after which they would hear 16 notes total before the next trial where this process would repeat. In each trial, each note was presented for 500 ms with a 1000 ms pause in between each note. Participants were notified when they had reach the end of each trial and each block, and after the completion of a block, had the option to take a short break. The hit rate, or accuracy of identifying the target tone, as well as the reaction time for press of the "T" key when the target tone was heard were both recorded for each participant and each trial.

When the participant had reached the end of the tenth block, they were notified, and were moved to a screen containing a full debriefing form. The participants then were instructed to email the experimenter with the method of delivery they preferred for their $10 compensation.

Results

All data analyses were done in Python (v. 3.11.4: d2340ef257) and R Studio (v. 4.3.1).

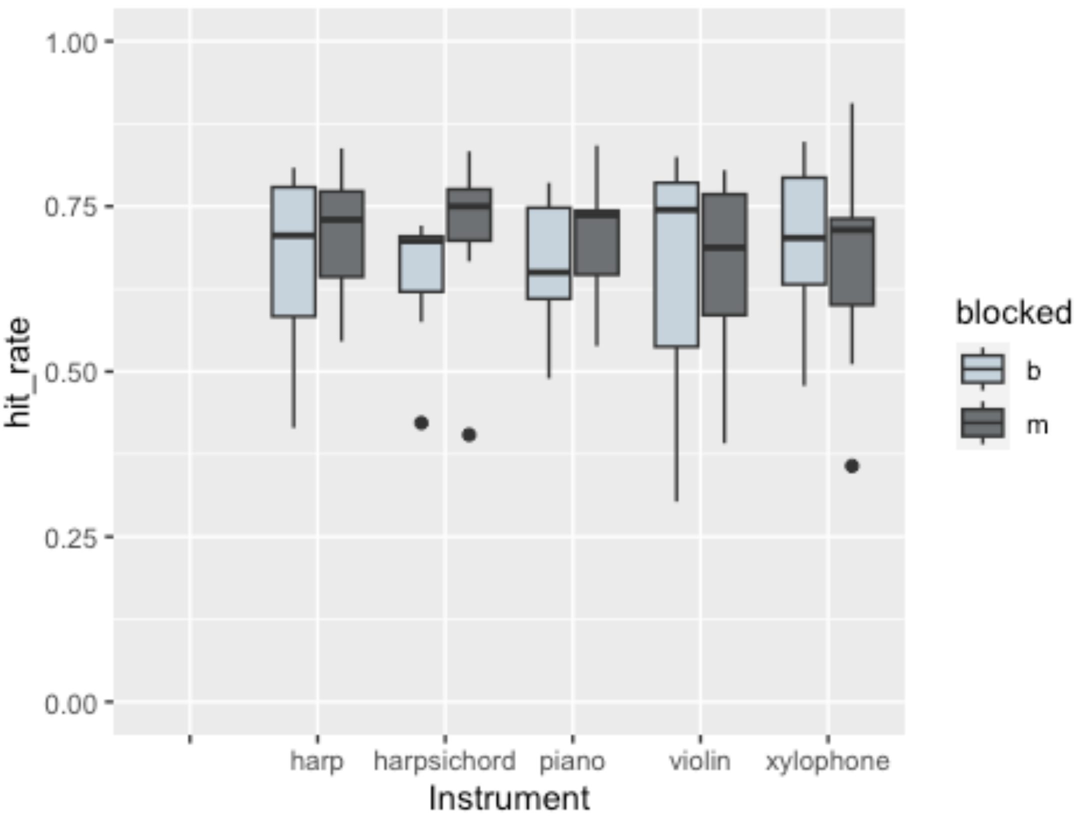
Figure 1. Hit rate by Instrument

Figure 1. shows the hit rate average by instrument for both blocked and mixed trials. In the graph, b stands for blocked, and m stands for mixed. This refers to whether or not a trial contained only one timbre or instead had a mixture of two.

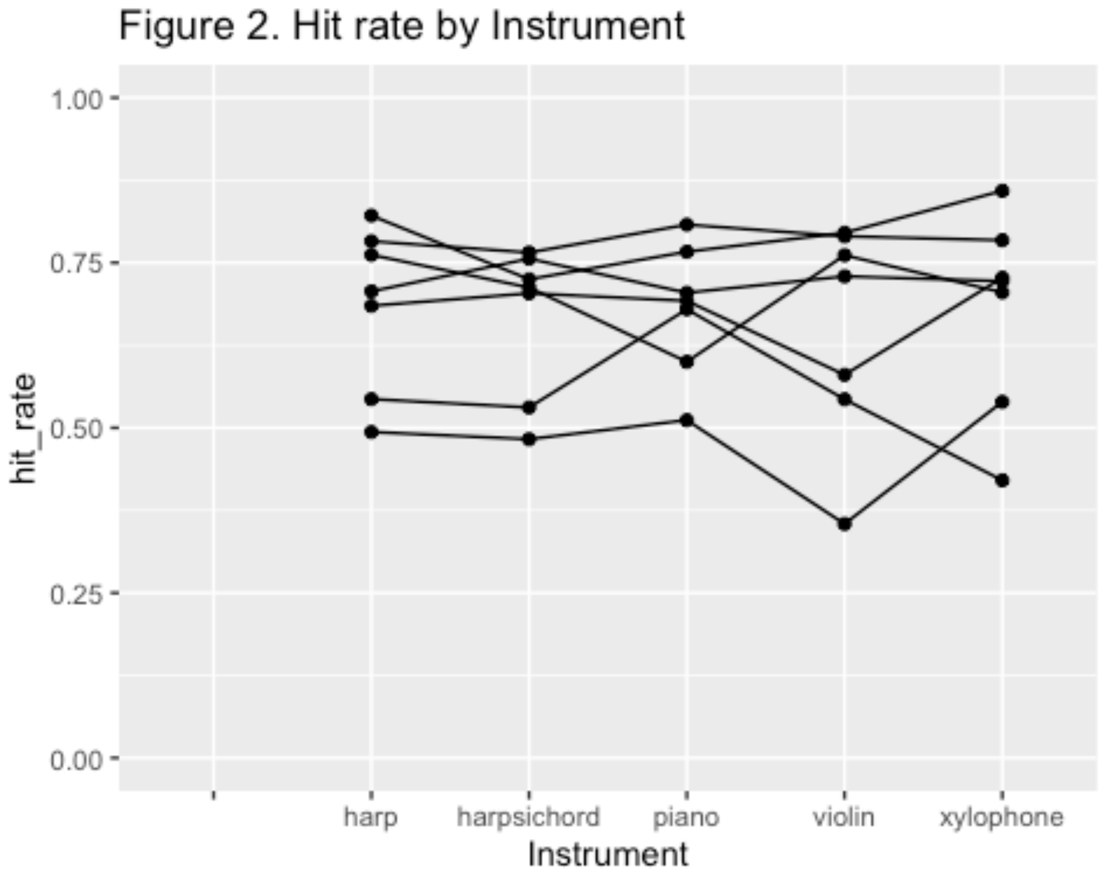

Figure 2. shows the same measurements as Figure 1. However, rather than show the average hit rate by timbre by trial type, it does so by participant. Each line represents a different participant's average hit rate per instrument.

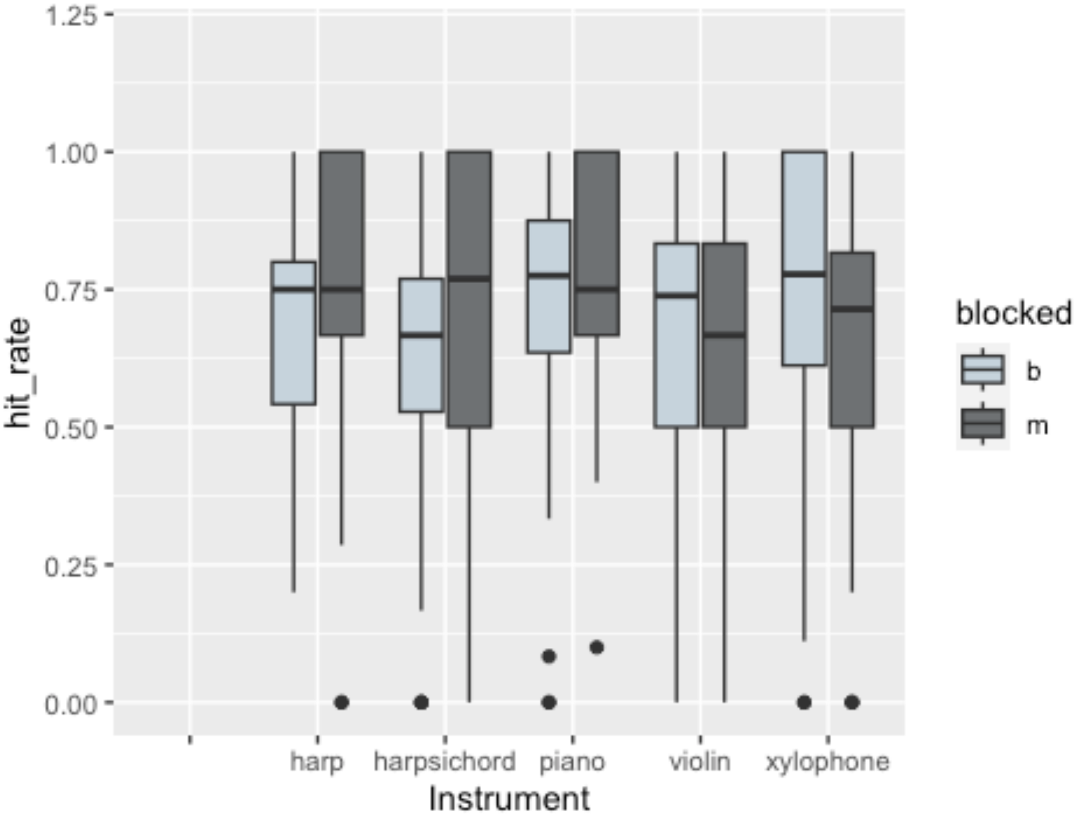

Figure 3. Hit rate by instrument timbre and target note

Figure 3. represents the average hit rate by instrument timbre and target note to be identified, where b stands for blocked trials and m stands for mixed. The mean hit rate for blocked trials was 0.69, whereas for mixed trials it was 0.71. A correlated samples t-test revealed that the difference in average difference in the mean hit rate by trial type was not statistically significant, *t* (6) = -1.08, *p* = 0.28. This means that regardless of trial type, there was no significant difference in average accuracy of detecting the target note by timbre.

Figure 4. Reaction times

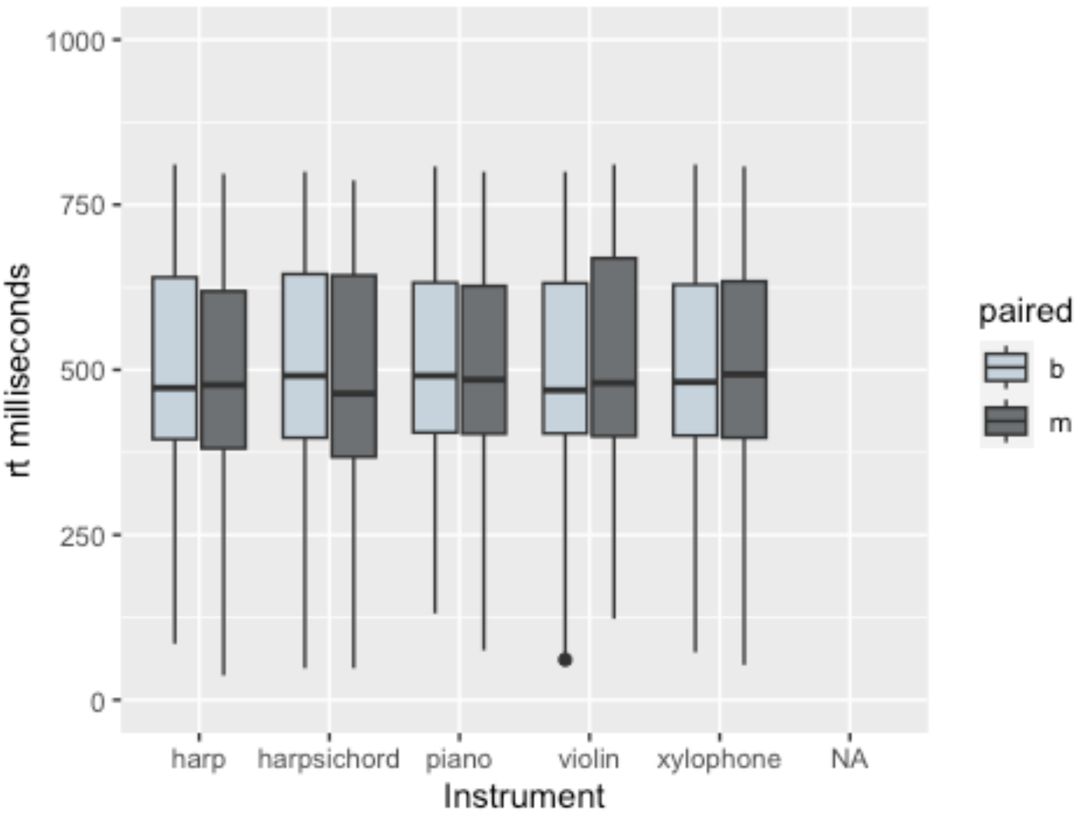

Figure 4. represents the average reaction times in milliseconds to the target note by instrument timbre in either the blocked or mixed trials, where b stands for blocked and m stands for mixed. The mean response time in milliseconds in blocked trials was 508.15 whereas for mixed trials it was 507.37 A correlated samples t-test revealed that the means between the two reaction times by trial type were not significantly different, *t* (6) = .12, *p* = .91 This means that regardless of trial type, there was no significant difference in the mean reaction times to the stimuli.

Discussion

The aim of this experiment was to take the paradigm used by Nusbaum and Morin (1992) and Magnuson and Nusbaum (2007) for processing costs indicated by greater reaction times when switching between two different talkers and instead apply this to the world of music with instruments of different timbres and a target note. Based on how much study has been done with musicians who possess AP, the goal here was to try and make an experiment that would have results more generalizable to all musicians by instead using the same types of experimental designs, but with musicians who did not possess AP.

Despite this goal, the initial hypothesis was not supported by the results of this experiment. There was no significant difference found in the average hit rates nor the average response times between blocked vs. mixed timbre trials when a target note was to be identified. The most logical explanation of these results based on the previous literature review rests in the fact that the sample size for this experiment was most likely not large enough to produce enough statistical power to find any significant differences between these two different trial types, as the originally hypothesized outcome has been found many times in AP individuals.

Another explanation for these results, though admittedly unlikely, is the possibility that this particular set of non-AP musicians was simply equally as fast and accurate at identifying the target note in the blocked timbre trials as they were in the mixed. Though again these results are most likely solely due to the small sample, the fact remains that findings from Van Hedger et al. (2016 & 2020) support the idea that AP musicians may indeed process pitch and have different memory for pitch than non-AP musicians. In this light, there is chance that some of the participants in this experiment had what Van Hedger et al. (2016) referred to as “pseudo-AP,” and thus simply had good enough pitch-memory that the timbre changing between trials in a mixed condition did not affect their performance to complete the task to the same degree as the blocked timbre trials. This would further support the idea that AP exists on a spectrum and is not a dichotomous discrete variable (Van Hedger et al., 2016).

However, as Van Hedger et al. (2020) also found, AP individuals may just have superior pitch processing memory to non-AP individuals, and as other studies have shown such as Van Hedger et al. (2015) that AP individuals do in fact show slower reaction times when having to identify a target note in a mixed rather than a blocked trial, this would work against the idea that this particular group of non-AP musicians would somehow have had superior pitch processing to AP musicians.

This study also included other variables such as a participant's familiarity with the experimental timbres based on the instrument they play, as well as musical expertise as measured by years of playing their instrument. Future analyses could be conducted with this data to see if the two aforementioned variables had significant interactions with hit rate accuracy or response time speed. Furthermore, this experiment should be replicated with a higher sample size to gain greater statistical power and see if results may differ from those found in this experiment, and instead be closer to the findings of Van Hedger et al. (2015).

To further the current scientific study of these processes, this experiment should also be replicated to instead use non-musicians. This kind of experiment could help establish a very generalizable baseline for how pitch and other auditory functions are processed in the brains of the average majority individual who listens to music but does not possess any formal musical instruction. As shown by Van Hedger et al. (2016) as well as a plethora of other studies, the average pitch memory for non-musicians may not be as good as those who have had instruction, but it is still quite accurate.

Hopefully the merit of this study can be extended further into the world of music cognition to more generalizable populations, and away from the more specific and limited study of AP musicians who have been featured so widely in this field. By testing other musicians as well as the general population using the paradigms in this experiment, we can help to increase our understanding of how pitch and other auditory stimuli are processed in the brains of a more general audience.

References

Bahr, N., Christensen, C. A., and Bahr, M. (2005). “Diversity of accuracy profiles for absolute pitch recognition,” Psychol. Music 33, 58–93.

Brammer, L. M. (1951). “Sensory cues in pitch judgment,” J. Exp. Psychol. 41, 336–340.

De Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. Behavior research methods, 47, 1-12.

Magnuson, J. S., and Nusbaum, H. C. (2007). “Acoustic differences, listener expectations, and the perceptual accommodation of talker variability,” J. Exp. Psychol.: Human Percept. Perform. 33, 391–409.

Nusbaum, H. C., and Morin, T. M. (1992). “Paying attention to differences among talkers,” in Speech Perception, Speech Production, and Linguistic Structure, edited by Y. Tohkura, Y. Sagisaka, and E. Vatikiotis-Bateson (IOS Press, Burke, VA), pp. 113–134.

https://pavlovia.org

R Core Team (2023). _R: A Language and Environment for Statistical Computing_. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.

The MathWorks Inc. (2022). MATLAB version: 9.13.0 (R2022b), Natick, Massachusetts: The MathWorks Inc. https://www.mathworks.com

Van Hedger, S. C., Heald, S. L. M., & Nusbaum, H. C. (2015). The effects of acoustic variability on absolute pitch categorization: Evidence of contextual tuning. *Journal of the Acoustical Society of America*, *138*(1), 436-446.

Van Hedger, S. C., Heald, S. L., & Nusbaum, H. C. (2016). What the [bleep]? Enhanced absolute pitch memory for a 1000 Hz sine tone. Cognition, 154, 139-150.

Van Hedger, S. C., Veillette, J., Heald, S. L., & Nusbaum, H. C. (2020). Revisiting discrete versus continuous models of human behavior: The case of absolute pitch. PLoS One, 15(12), e0244308.

Van Rossum, G., & Drake, F. L. (2009). *Python 3 Reference Manual*. Scotts Valley, CA: CreateSpace.

Acknowledgments

Matthew C. Di Santo, Master of Arts Program in the Social Sciences with a concentration in Psychology degree candidate (The University of Chicago).

I would like to first and foremost thank Letitia Ho for her instrumental help in this experiment by graciously lending me her code for the experiment itself and making the substitutions needed to match my paradigm. I would also like to thank her for lending me her analysis codes so that I could change them to what worked for me to analyze my results. I would next like to thank Dr. Stephen Van Hedger for lending me his stimuli from his cited 2015 paper, as well as making new stimuli for me to use in this particular experiment. I would like to also thank Dr. Howard Nusbaum, Dr. Shannon Heald, and Jean Matelski-Boulware for giving me feedback and guidance on how to move forward with my thesis and experiment.

Outside of the lab, I want to especially thank Dr. Tori Gross, Dr. Danielle Bolling, and Dr. Samantha Fan for all giving me feedback and guidance towards the best steps to help me finish this degree, as well as giving feedback on my writing. I would also like to thank E.G. Enbar for always keeping me on top of deadlines and making sure I had all of the information I needed to graduate.

Lastly I would like to thank my friends and family, as without their help and support, I could not have finished this degree.