Spring 2023

A Comparative Investigation of Tonal Memory Improvements with Electronic and Vocal Pitch Stimulus Training

Grayson M. Fletcher
University of South Carolina - Columbia

A Comparative Investigation of Tonal Memory Improvements with Electronic and Vocal Pitch Stimulus Training

By

Grayson Fletcher

Submitted in Partial Fulfillment
of the Requirements for
Graduation with Honors from the
South Carolina Honors College

April 2023

Approved:

Jessica Green
Director of Thesis

Elizabeth Myers
Second Reader

Steve Lynn, Dean
For South Carolina Honors College

ACKNOWLEDGEMENTS

I would first like to thank Dr. Jessica Green for directing me on this project and for her significant contributions to the construction of the PsychoPy training program, the experimental setup, and extraction of data.

I would also like to thank Ms. Elizabeth Myers for guidance and revisions on technical aspects of academic writing and reporting of research.

Great thanks to the faculty and staff of the Honors College at the University of South Carolina for being a consistent resource not only in the course of this work but over all 4 years of my undergraduate studies.

ABSTRACT

The study described herein compares improvements in tonal memory for young adults (age 19-22), specifically retention of pitches, between two groups (n_1=8, n_2=8) with nearly identical training programs. One group was provided electronic pitch stimuli for the training program, while the other group was given human vocal stimuli. Self-paced computerized training sessions were conducted in a soundproofed testing room and interposed between pre-test and post-test measurements. A pre-test—post-test randomized experimental design allowed for assessment of whether the training was effective for each group in addition to comparing effectiveness of training between groups. Analyses with demographic factors, particularly previous musical experience and experience with tonal languages, are also discussed. No significant improvements were found over the course of the training for either group, with averages on pre-tests and post-tests falling between 35% and 36% accuracy in all cases. The hypothesis that the vocal pitch stimulus group would experience greater improvements than the electronic stimulus group is not supported. No correlation between previous musical experience and baseline tonal memory assessment was found. Limitations of this research include small sample size leading to high variability in addition to the brevity of the computerized training program.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION AND LITERATURE REVIEW

Absolute pitch (AP), commonly referred to as perfect pitch, is a quality present in an estimated 1 in 10,000 of the general population (Takeuchi & Hulse, 1993). Individuals with AP possess the ability to identify and name the musical pitch, for example, C4, of a tone without reference. Furthermore, most AP possessors can vocally produce a named pitch on command consistently, with little to no deviation from the absolute pitch center. The factors contributing to AP remain unknown, however various studies have proposed that factors such as early exposure to music, natively speaking tonal languages such as Mandarin Chinese, and genetics may play a role. For example, a 2004 study at the University of California investigated native speakers of Mandarin and Vietnamese, languages for which tonality is a significant component in the meaning of a word. A highly precise stability in pitch when enunciating specific words over several different days was demonstrated for both the Mandarin and Vietnamese speakers, while this stability was lacking for English speakers, suggesting that pitches may become associated with their verbal labels during acquisition of speech during childhood for speakers of tonal languages (Deutsch, Henthorn, & Dolson, 2004). This hypothesis is further supported by a higher occurrence of AP in native tonal language speaking populations (Moulton, 2014).

Numerous efforts with varied approaches have failed to develop perfect pitch in adults. For this reason, the notion that AP is developed during a critical period in early childhood for those with some rare genetic makeup has long been accepted. It is hypothesized that, like language acquisition, the ability to label and subconsciously remember pitches must be developed during this critical learning period, or else it is purely genetic. However, a recent study from The Psychonomic Society, Inc. at the Chinese University of Hong Kong challenges this notion, claiming that previous studies had insufficient sample sizes and inadequate AP training. Contrary to the findings of their predecessors, 14% of participants in the study, who did not previously exhibit AP, were able to name 12 pitches with an accuracy of 90% or higher after a 12-40 hour adaptive, personalized, and computerized training program (Wong et al., 2020). This figure of 90% accuracy is comparable to the accepted definition of AP, suggesting that it may be a quality that is learnable in adulthood.

The prospect of a connection between tonal language acquisition and AP raises the question of whether the auditory qualities of human vocals play a role separate from linguistics in tonal memory. Rather than attempting to develop AP, this study will focus on the extent to which tonal memory can be developed after childhood with a specific computerized training program. The proposition is that although true perfect pitch may not be attainable for development in adulthood, tonal memory can be improved through practice. Specifically, it is hypothesized that individuals in early adulthood will have greater improvements in tonal memory with vocal pitch recognition training than a group with electronic pitch recognition training. Although AP itself may not be learnable, perhaps vocal pitch recognition differs from the mechanism of identifying pure tones, and could present a route to a degree of ability indistinguishable from individuals with AP. While the debate over the learnability of true AP during adulthood is still open, if individuals show improvement in accuracy for the memory of pitches after a specific training program, this may illuminate a mechanism of tonal memory with potential future applications to AP development in adulthood. This would not only serve to inform future tonal memory training methods, but it could also provide behavioral evidence in support of one proposed neuroscientific mechanism or another for tonal memory development.

More specifically, a 2017 study has suggested that some subconscious order of absolute pitches may exist in the general population, not just AP possessors (Eitan, Ben-Haim, &

Margulis, 2017). This study could provide direction for a method of bringing this subconscious framework into a conscious understanding of labeled pitches by indicating whether or not vocality plays an important role in the utility of such a framework. Additionally, in an investigation of individuals with amusia (impaired musical perception), it was found that although tonal sequences were not encoded into their working memory, some implicit processing of tonal structures was preserved even for amusics (Albouy et al., 2013). This motivates an argument that perhaps it is not individual pitches that are remembered, but rather tonal frameworks (i.e. the key of G major) that are implicit in memory and used to label a note for individuals with AP.

The anticipated results were that both the vocal pitch memory and electronic pitch memory training groups would demonstrate significant improvements in measured tonal memory after their respective training sessions. Furthermore, it was expected that the vocal pitch memory training group would exhibit greater improvements in tonal memory than the electronic pitch group, following from an argument of greater mental rehersability of vocal pitches in comparison to electronic pitches. Researchers from the Auditory Cognition and Psychoacoustics Team at the Lyon Neuroscience Research Center and the Developmental Cognitive Neuroscience Unit at the UCL Institute of Child Health investigated the mechanisms of auditory working memory by using three distinct auditory stimuli: words, tones, and timbres. The study found that working memory processes for timbres were different than those for words and tones. The memory of words and tones seemed to depend on so-called "rehearsal mechanisms," however the sensorimotor codes involved in the two stimuli are likely different. These results imply that auditory working memory is an involved, non-unitary system. It is suggested that while words and tones can be actively rehearsed in working memory, memory of timbres cannot be facilitated by the same type of rehearsal processes. Between electronic pitches and vocally produced pitches, the present study proposes that the qualities of the human voice could enhance the rehersability of pitches, leading to the stated hypothesis.

Additionally, improvements in tonal memory were expected to vary inversely with amount of previous musical experience, as the baseline is anticipated to be higher due to existing aural skills and any previous pitch training, such as call and response. Since extensive training in a single subject area tends to follow a typical learning curve, it is possible that those with a lot of musical experience have developed good relative pitch and may experience no improvement over a single training session.

These results would firstly provide further evidence that tonal memory can be improved through computerized training methods. Combined with the existing evidence that some subconscious absolute framework of pitches exists even in individuals without AP, these findings would indicate potential methods of fine-tuning this framework and bringing it to conscious access. Additionally, this would serve to support that kinesthetic learning methods (in this case, associating muscle memory of vocal production of a pitch with the absolute pitch label) may be effective for development of tonal memory, for which there is existing evidence in studies of the use of solfege hand signs and instrument fingering patterns to improve sight-singing (Frey-Clark, 2017; Howland, 2020). Ultimately, this study is designed not to provide evidence of AP development in early adulthood, but rather to evaluate a specific method of improving tonal memory towards a functionality similar to AP. Since the neural mechanisms underlying the development of AP are unknown, no current behavioral study would be sufficient to conclude that development of the functionality of AP is mechanistically the same as that of "natural" AP possessors. Rather, this study is meant to work within the framework of existing literature to point in the direction of a mechanism by which absolute pitches can be better remembered and consciously labeled in adulthood.

METHODOLOGY

As examined earlier, previous studies have suggested that tonal memory may be improved by computerized training (Wong et al., 2020). The current research has employed a pre-test—post-test design to compare improvements in tonal memory of young adults (age 19-22) between a group with electronic training stimuli (n=8) and a group with human vocal training stimuli (n=8). The independent variable was the type of training stimulus (pure electronic tone or vocally produced tone), and the dependent variable was tonal memory, measured by the pre-test and post-test. To control for confounding environmental factors, all training sessions were completed in the same soundproofed testing room using the same audio source. The pre-test, training program, and post-test were constructed and administered using PsychoPy.

To obtain data, I recruited a sample of 16 college students at the University of South Carolina (median age=22), randomly assigning each student to either the electronic pitch group or the vocal pitch group. Each subject, after establishing their consent to participate, completed a short survey regarding previous musical experience, experience with various tonal languages, and other demographic information, to identify possible factors or sources of confounding in data analysis. Participants were asked not to hum, sing, whistle, or use a tuner to aid them in remembering pitches, then they were given the pre-test, followed by 7 training modules and the post-test. All tasks were identical between the electronic group and the vocal group besides the source of the auditory stimuli. One minor exception to this involving the pre-test will be discussed later.

The pre-test and post-test identically consisted of a tone identification task. Using a computer program, subjects were first provided with an explore page of 12 buttons labeled

corresponding to the 12 most common pitches in the Western standard equal temperament tuning system, in ascending order from left to right. Tasked with remembering all 12 pitches, participants were encouraged to spend several minutes using the buttons to get familiar with the tones in any way they thought may be effective. Proceeding immediately to the testing task, participants were provided an auditory stimulus of one of the 12 pitches and asked to identify which pitch they heard using identical buttons to the explore page. This was repeated for all 12 pitches in random order without access to the explore page in between stimuli. The pre-test was used to establish a baseline for each participant's ability to remember pitches, and the post-test to measure the effectiveness of the training program. Performance on the pre-test and the post-test was assessed by two measures: (1) the number of correctly identified pitches, and (2) the average absolute interval between the correct response and the actual response (for example, if the correct response was A, but the actual response was A#, the difference was recorded as 1).
The pre-test and post-test were designed to be the most challenging tasks in the experiment such that the baseline was not too large and improvements could be detected. On the other side of this issue of sensitivity, a test too challenging runs the risk of becoming random and losing all sensitivity to improvements. Implications of this trade-off are further discussed with relevance to the findings in the analysis and conclusion sections.

Another potential issue with the pre-test and post-test is that the 12 pitches included for the electronic stimulus group were not the exact same pitches as for the vocal training group (9 pitches were shared between the groups), though both cases included 12 consecutive semitones in ascending order. This was not an intentional feature in the experiment's design, and possible effects on the results of the study will be addressed.

The computerized training program consisted of 7 different tasks/modules and was divided into 2 major sections (tasks 1-4 and tasks 5-7). Within each section, tasks were intended to increase in difficulty, such that tasks 4 and 7 were the most difficult and tasks 1 and 5 the easiest, with all participants completing tasks 1-7 in numerical order. Additionally, each of the 7 training tasks included first a section with feedback and then a section without feedback. This choice of design allowed for intermittent assessment of each participant's improvements in individual tasks throughout the training session in addition to the final evaluation. Though the training was self-paced, it was designed to be completed in approximately 45 minutes.

I will describe each of the training tasks and its purpose in detail. The first 4 tasks focused on perceptual matching and relative pitch. Module 1 consisted of a single sample tone, followed by a single test tone, and participants were asked to indicate whether the tones were the same or different. This task was primarily meant as a "warm-up" for subjects to get familiar with the structure of the experiment and listen closely to the stimuli. In module 2, a sample sequence of 3 pitches was followed by a single test tone, with the question of presence or absence of the test tone in the sample sequence. This required subjects to remember all 3 pitches in the sample sequence and then compare them to a 4^{th} pitch as a perceptual matching task. In module 3, a 3-tone sample sequence was again followed by a single test tone, with the task of identifying the location (or absence) of the test tone in the sample sequence. This expanded on the presence/absence task by requiring subjects to also remember the order of pitches in the sample sequence. Module 4 consisted of two sequences of 3 tones, and participants were asked whether the two sequences were identical, contained the same pitches in a different order, or contained at least one different pitch. This task was the most involved perceptual matching task, requiring separate memory of two sequences and comparison of each pitch in both sequences.

The last 3 tasks focused on identification of pitches by name, more similar to the pre-test and post-test. Module 5 began with an explore page of 3 possible tones, and participants had to identify which of the 3 tones was played for a series of single test pitches. This task was intended as an introduction to the identification portion of the training, similar to task 1. Module 6 included an explore page of 6 possible tones, and subjects had to identify a series of single test pitches given a choice of 3 options. Finally, module 7 was identical to module 6 except that all 12 tones were possible, but participants still only had to pick from 3 provided answer choices. Modules 6 and 7 motivated subjects to construct a mental framework of pitches using the explore page, then use their memory to pick out the test pitch. Module 7 concluded the training portion of the experiment and was followed immediately by the post-test.

Additionally, throughout the 7 training modules, the elapsed time between the sample pitch or sequence and the test pitch was a discrete variable of either 1, 3, 6, or 9 seconds, which occurred equally often and randomly. This variable introduced another dimension designed to train memory over longer intervals of time. It also allowed for analysis of improvements in tonal memory for each individual task.

RESULTS AND DISCUSSION

Of the 16 college students in the sample, 62.5% had some level of previous musical experience, and none had experience with tonal languages. Across both groups, the average pre-test score was 36.5% correct with an average interval of 1.36 semitones from each correct response, and the average post-test score was 35.9% with an average interval of 1.22 semitones. Individual pre-test and post-test data for each group are given in Tables 1 and 2 below, sorted by years of musical experience.

Table 1: Pre-Test and Post-Test Results for Electronic Pitch Stimulus Group

		Pre-Test		Post-Test		Difference	
Participant Number	Music Experience (years)	Accuracy	Average ME	Accuracy	Average ME	Accuracy	Average ME
007	0	0.42	0.83	0.33	1.42	-0.08	0.58
009	0	0.33	1.42	0.42	1.33	0.08	-0.08
011	0	0.42	0.67	0.25	1.08	-0.17	0.42
015	0	0.17	2.58	0.17	2.00	0.00	-0.58
001	3	0.25	1.33	0.50	0.58	0.25	-0.75
013	4	0.25	1.75	0.25	1.25	0.00	-0.50
003	6	0.33	1.08	0.33	1.08	0.00	0.00
005	10	0.75	0.75	0.67	1.17	-0.08	0.42
Average	**2.88**	**0.36**	**1.30**	**0.36**	**1.24**	**0.00**	**-0.06**
Standard Deviation	NR	0.18	0.64	0.16	0.40	0.13	0.51

Note: ME represents the average margin of error, in semitones, of each participant’s responses on the pre-test. ME is not a figure of statistical error in the measurement. Tabulated values will retain only two decimal places to reduce clutter. See appendix A for access to high precision values.

Table 2: Pre-Test and Post-Test Results for Vocal Pitch Stimulus Group

		Pre-Test		Post-Test		Difference	
Participant Number	Music Experience (years)	Accuracy	Average ME	Accuracy	Average ME	Accuracy	Average ME
008	0	0.25	2.25	0.50	0.92	0.25	-1.33
016	0	0.67	0.67	0.25	1.58	-0.42	0.92
004	1	0.58	0.58	0.25	1.17	-0.33	0.58
012	1	0.08	1.75	0.25	1.58	0.17	-0.17
006	2	0.25	1.67	0.50	1.67	0.25	0.00
018	3	0.58	0.50	0.33	1.25	-0.25	0.75
014	7	0.25	2.92	0.33	0.67	0.08	-2.25
010	8	0.25	1.08	0.42	0.75	0.17	-0.33
Average	**2.75**	**0.36**	**1.43**	**0.35**	**1.20**	**-0.01**	**-0.23**
Standard Deviation	NR	0.21	0.87	0.11	0.39	0.28	1.09

In Tables 1 and 2, the accuracy refers to the proportion of correctly identified pitches and is provided for both the pre-test and the post-test, while the ME represents the average interval between the participant's response and the correct response. Because the test was designed to be very difficult and consisted of only 12 questions, the ME is a more comprehensive measure of a subject's performance.

From Table 1, it is first notable that for the group which received training with electronic stimuli, the average accuracies were identical, at 36%, between the pre-test and the post-test. This is quite surprising given the small sample size of n=8. The standard deviation of accuracy differences between the pre-test and post-test was 13%, reflecting the high level of variability for a small sample. The average difference in ME was -0.06, meaning that participants responded 0.06 semitones closer to the correct response on average in the post-test than in the pre-test. The results of a paired t-test confirm that this figure does not represent a statistically significant improvement in performance at the 95% confidence level ($t= -0.347$, $p=0.37$).

In Table 2, the average accuracies for the vocal pitch group were again nearly identical between the pre-test and post-test, with a difference of only 1%. This difference represents the weight of a single response by a participant on either the pre-test or post-test. The standard deviation in accuracy differences is very large, at 28%. It is interesting that despite high levels of variability and a small sample size, differences in the mean test score are effectively zero. Possible explanations for this finding will be later discussed. The average difference in ME, however, was -0.23, representing a marginal improvement. The standard deviation in ME differences was 1.09, making the improvements insignificant by analysis in a paired t-test ($t= -0.596$, $p=0.29$).

Performances on the pre-test and post-test for the electronic group and vocal group are represented graphically in Figures 1 and 2, respectively.

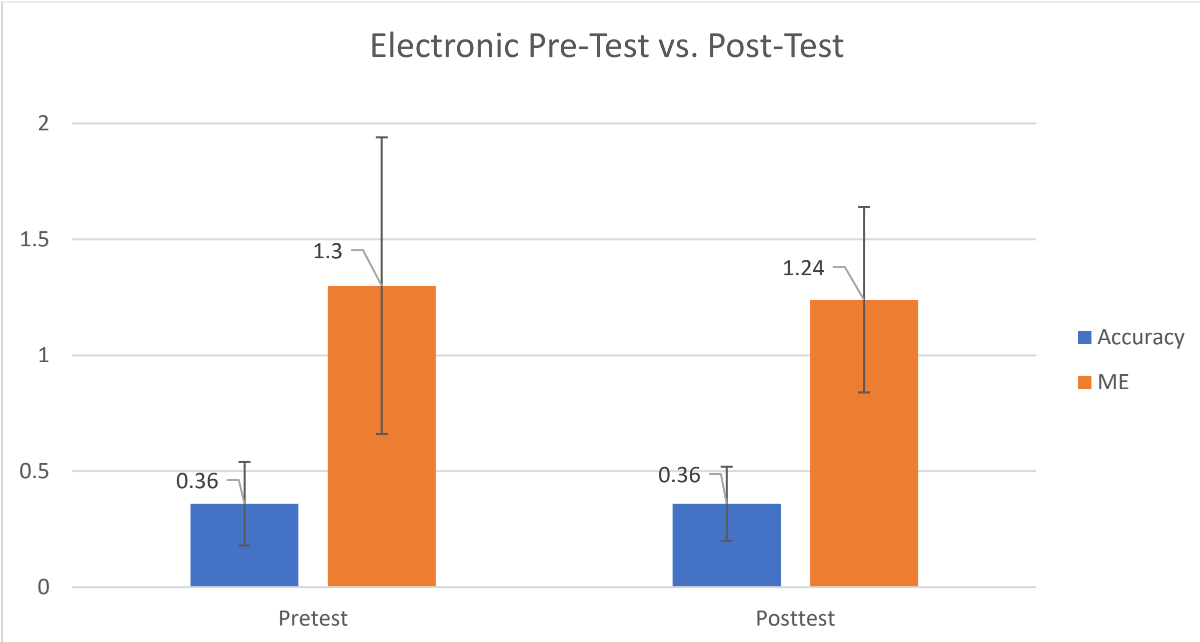


Figure 1. Comparison of pre-test and post-test performance for electronic group. Error bars represent standard deviations.

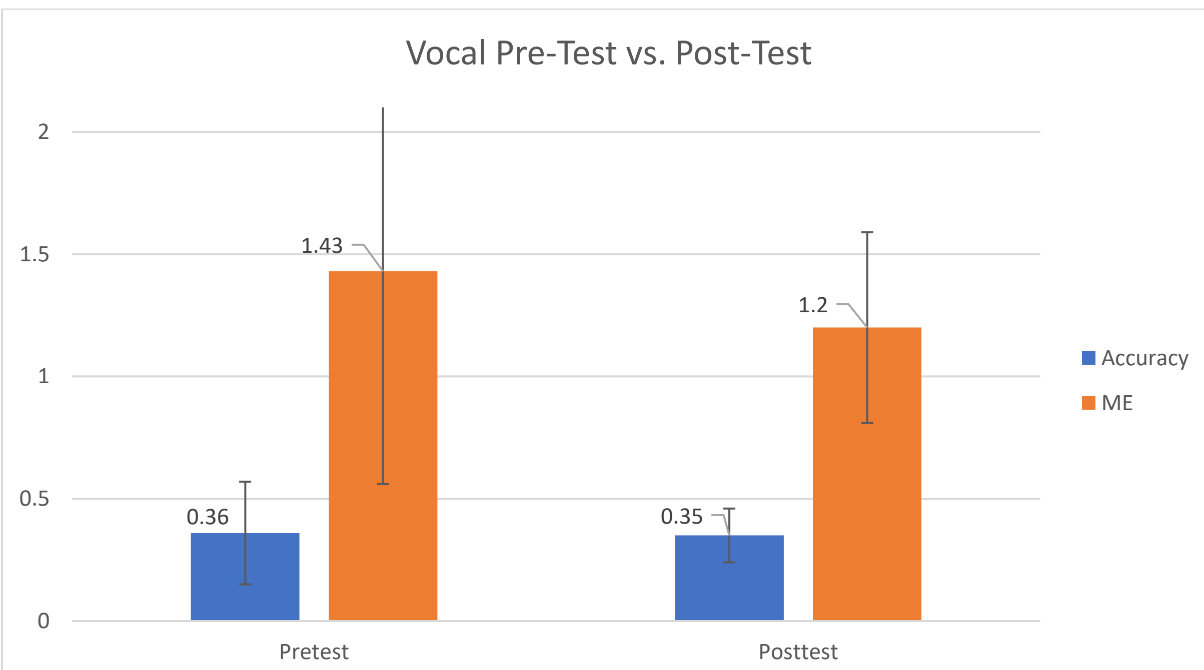


Figure 2. Comparison of pre-test and post-test performance for vocal group.

In a preliminary comparison of Tables 1 and 2, it stands out that the baseline accuracy scores were identical between the electronic stimulus group and the vocal stimulus group.

Neither group exhibited any noticeable increase in accuracy. Comparing improvements in the average margin of error, the vocal group improved by an average of 0.17 semitones more than the electronic group did. Results of a 2-sample t-test determine that the vocal improvements in margin of error are not significantly greater than those in the electronic group ($t= -0.397$, $p=0.35$).

It is clearly apparent by comparison of Figures 1 and 2 that the tonal memory training was not effective to improve performance on the pitch identification test for either group, nor was there any significant difference between any measure of performance between the two groups on either the pre-test or the post-test. A major limitation of the scope of these conclusions is presented by the large variance in performance. There are two sources of this large variance: (1) the small sample size, and (2) characteristics of the pitch identification test. Since the test only included 12 data points, each response was responsible for 8.3% of an individual's accuracy score. Therefore, a single mouse slip on either the pre-test or the post test would affect the entire group average improvement by 1%. A second feature of the test that would further contribute to high variability is its degree of difficulty. Because the test was designed for participants to respond incorrectly more often than correctly, a few lucky guesses would significantly help a participant's accuracy. For this reason, the margin of error was expected to be a better measure of performance. However, comparing the relative standard deviations of ME to those of accuracy within each group, neither measure seems to be more precise than the other.

It was expected that individuals with more previous musical experience would perform better on baseline assessments for both groups. Scatterplots of accuracy against years of musical experience are given for each group in Figure 3, and ME is plotted against musical experience in Figure 4.

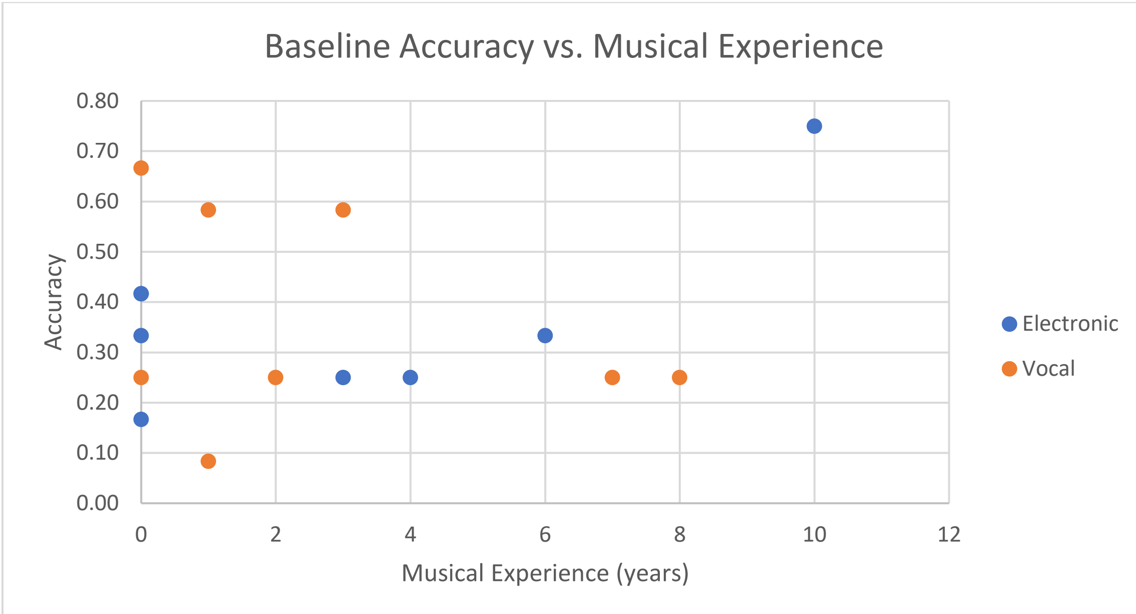


Figure 3. Accuracy performance on the pre-test by years of musical experience.

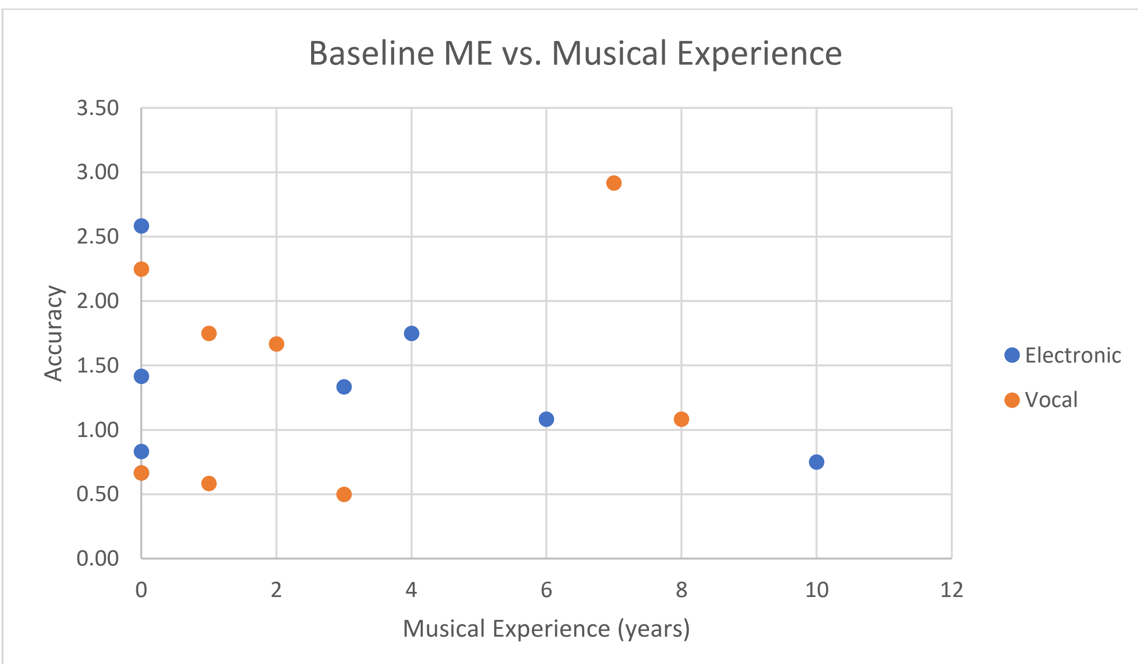


Figure 4. Average margin of error on the pre-test by years of musical experience.

No correlation between baseline performance and musical experience is found in Figures 3 and 4, contrary to expectations. Within the electronic pitch group, there was one outlier with an accuracy of 75% on the pre-test. This participant reported 10 years of musical experience, suggesting that perhaps there is a steep learning curve over years of experience for tonal memory. It was verified that this participant is not an AP possessor but does have significant training in relative pitch.

Then, it would be expected that improvements in performance over the training session would also not correlate with musical experience, since the reason for the original hypothesis that individuals with musical training would have smaller improvements was the expectation of a higher baseline, giving less room for growth. Individual differences in performance are plotted against musical experience for the two groups in Figures 5 and 6, confirming that there exists no correlation.

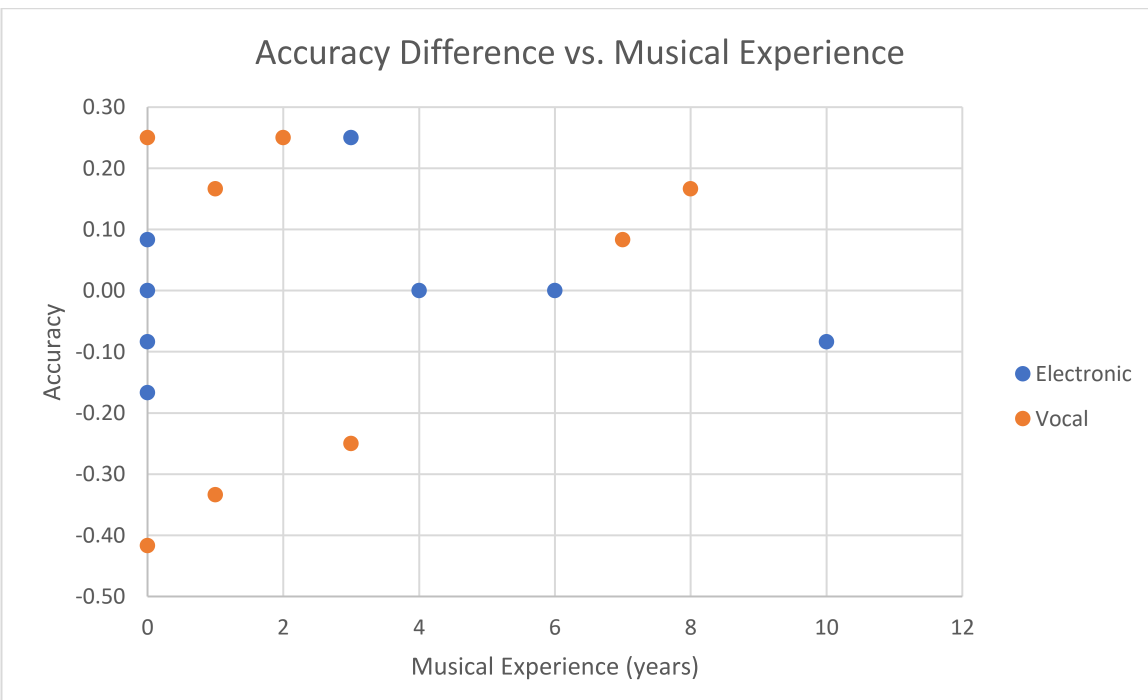


Figure 5. Differences in accuracy between the pre-test and post-test versus musical performance.

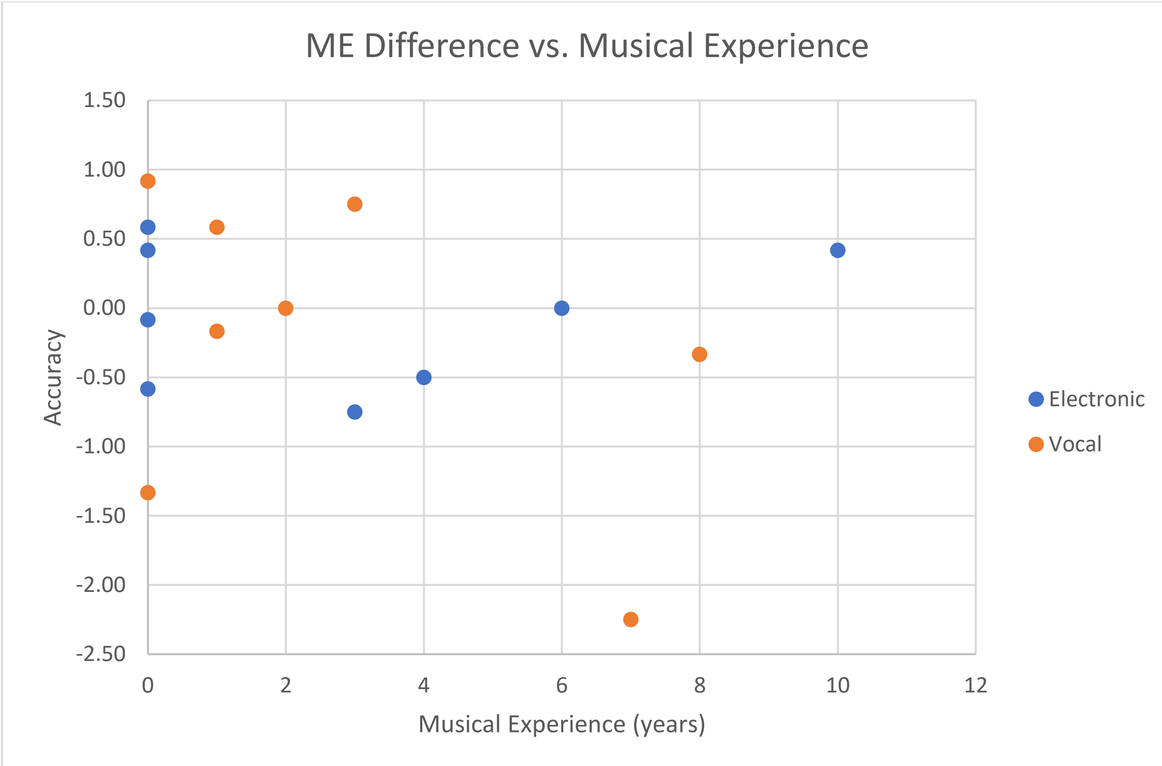


Figure 6. Differences in margin of error between the pre-test and post-test versus musical performance.

Thus far, none of the three major hypotheses set forth in the motivation of this study have been supported. Neither group improved in a pitch identification test over the course of the training, the vocal stimulus group did not exhibit greater improvements than the electronic stimulus group, and no correlation was found between years of musical experience and performance.

Returning to the question of why average accuracy scores were nearly identical for all 4 sets of the pitch identification test (pre- and post- test for each group), one possible explanation relies on a conjectural assumption, but it is primarily statistical. Because the test involved picking out a single pitch from a discrete spectrum of 12 pitches, it would be reasonable for participants to mentally divide the spectrum into 3 groups – low, middle, and high. Then, a random guess within each group would have a success probability of 25%, with a standard deviation of 1.5 correct responses (12.5%) over 12 trials. The most common score on the pre-test and the post-test was 25%, giving some support for this explanation, while the average was 35%-36%, which would indicate that participants performed better than chance in guesses narrowed down to 4-note groups of the spectrum. While the mode was 25%, at least one participant performed at 50% or above in every set, contributing to higher variances than expected for the case of random trials of 25% success probability. A plausible explanation for why some participants performed with such accuracy could be due to the randomization of the order of pitch stimuli in the test. In most cases, randomization is necessary to the validity of an experiment, however, in the case of an order of pitches, it is possible that some participants had an "easier" test than others if stimulus pitches with small intervals between them appeared in successive test questions.

Before concluding the discussion and addressing the question of why the tonal memory training programs were not effective for either group, it is important, for the sake of thoroughness, to consider the intermittent evaluations within individual training tasks. The average proportion of correct responses on each task for each group are tabulated on the following page.

Table 3: Average Performance by Training Task for Electronic and Vocal Stimulus Groups

	Electronic		Vocal	
Task	Feedback	No Feedback	Feedback	No Feedback
1	0.95	0.96	0.97	0.98
2	0.69	0.75	0.70	0.74
3	0.66	0.59	0.66	0.66
4	0.68	0.67	0.72	0.74
5	0.83	0.95	0.95	0.95
6	0.71	0.65	0.72	0.68
7	0.66	0.67	0.66	0.64

As expected, performance was highest for tasks 1 and 5, designed to be introductions to the two major sections of the training. Accuracy decreased in order of tasks 1 to 3 for both groups, and improvements over the course of a task (feedback to no feedback) occurred in both groups for task 2. Surprisingly, the vocal group only exhibited improvements in tasks 1, 2, and 4, while the electronic group improved in tasks 1, 2, 5, and 7. Although no improvements were statistically significant at the 5% significance level, the increase in accuracy from 83% to 95% for the electronic group in task 5 appears to be notable (paired t-test p=0.10). The average performance with feedback in task 5 was lower than expected, as participants had only 3 notes (low, middle, and high) to choose from, however this level of inaccuracy can be sourced to an outlying data point of 33% accuracy, while all other marks were 75% or higher. Similarly to performances on the pre-test and post-test, there is no evidence within training task averages to suggest effectiveness of either pitch stimulus training program.

Further analysis of the training tasks can be conducted by subdividing exercises by ISI (the time elapsed between the stimulus sequence and the test sequence). It was expected that participants would perform worse on tasks that required longer memory of a pitch or sequence. The results of this analysis yielded a random distribution of performance with respect to ISI. I have chosen not to include these figures in the text to avoid prolonging attention to non-correlated variables.

CONCLUSION AND FUTURE DIRECTIONS

Considering these results, the theory that vocal pitch stimulus training would be more effective than training with electronic stimuli is not supported. Several factors likely contributed to the general ineffectiveness of the tonal memory training described in the methodology. Most significant is the brevity of the training. Tasks were simplified and the training was shortened to a length of 45 minutes for the purposes of designing a feasible computer program and recruiting more willing participants. The Chinese University of Hong Kong study discussed in the introduction, which saw significant improvements in pitch identification, used a personalized and adaptive computerized training of 12-40 hours, which was much more extensive than the program used here (Wong et al., 2020). Additionally, returning to the earlier mentioned relationship between test difficulty and sensitivity to improvements, the consistency of performance of only 35%-36% across all instances of the primary measurement suggests that the threshold of difficulty may have been exceeded such that participants could have been guessing within a range of the possible tones on the pre-test and post-test.

Another possible explanation for a lack of measured improvements on the post-test is that participants were simply fatigued after completing the pre-test and training in the same session. After 45 minutes of hearing the same tones and completing various memory tasks, exhaustion may have impacted post-test performances. Nevertheless, this research finds no evidence for the argument of greater mental rehearsability of vocal pitches in comparison to electronic ones.

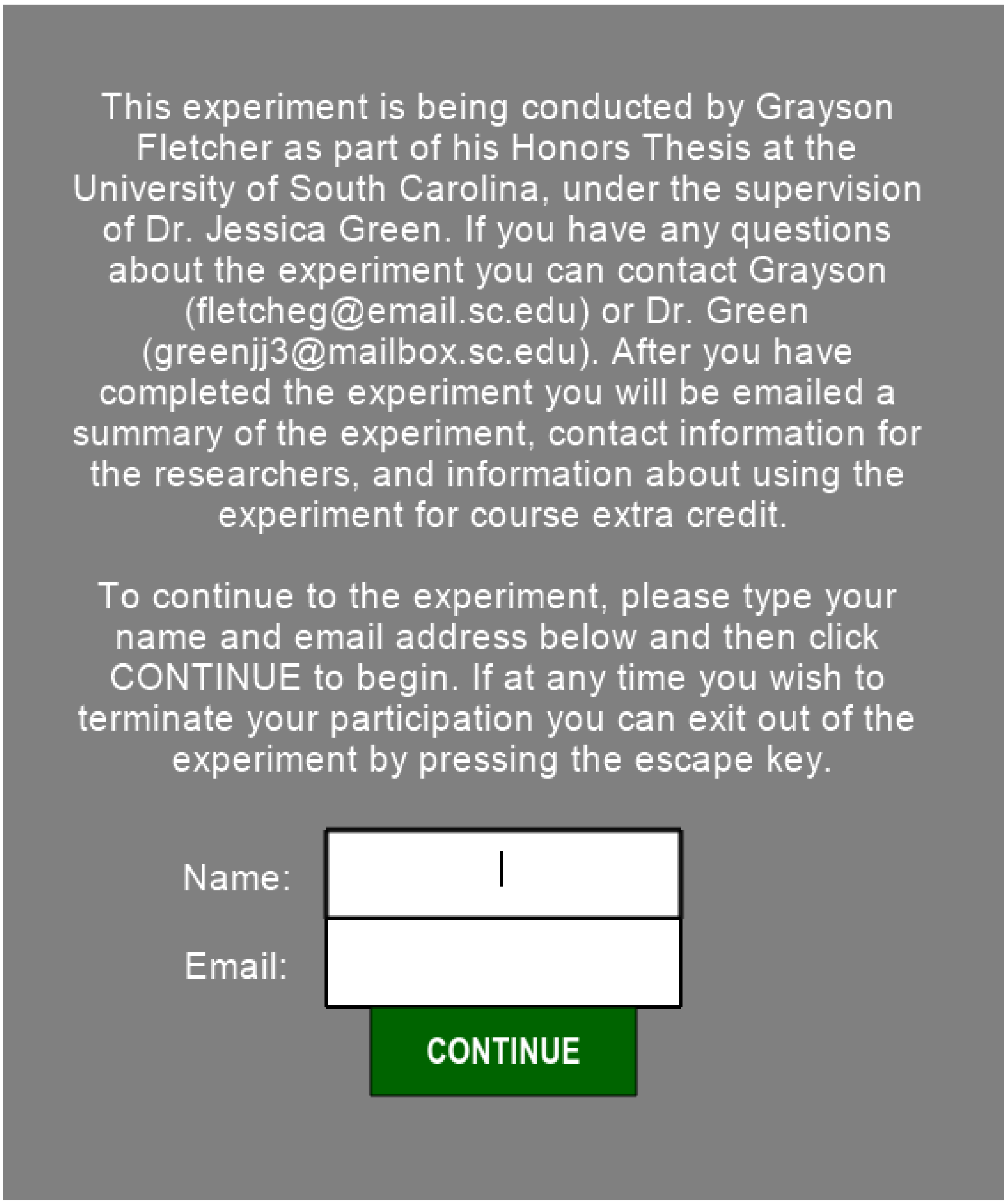
One limitation of these conclusions is the high standard deviations of test measurements. Relative standard deviations were as high as 50% for accuracy and 49% for average ME, meaning that average scores could be halved and still fall within one standard deviation. The primary contribution to large standard deviations was the small sample size, though they directly represent inconsistencies between individual performances on the test. A second limitation of the conclusions, again stemming from the high variability, is the lack of sufficient statistical power to detect differences in measurements. With the same standard deviations, accuracy improvements of 11% in the electronic group and 23% in the vocal group would have been required to conclude statistical significance.

A limitation of the study is that the pre-test and post-test did not vary length of time in a measurable way between hearing a known pitch and then being asked to identify it. Participants were free to explore the 12 pitches before starting the test, however they did not have access to this reference at any time between the 12 test questions. Perhaps allowing participants to access the explore page in between each question, then varying the length of waiting time before receiving a test pitch, would provide a better measure of tonal memory. This was considered in the planning phase of the experiment, but it would have significantly increased the time required to participate and was deemed too tedious. Although the experiment was conducted in a controlled, soundproofed environment, the original intention was to have participants complete multiple online training sessions in their own time. Difficulties with compatibilities of the PsychoPy online platform led to the in-person setup, limiting the number of participants and the extent of the training. Another limitation of the computerized design of the study is that the vocal stimuli were recordings of human vocals, and not live vocals. Thus, the vocal stimulus group received electronic simulations of human vocals which, for the most part, have an identical wave structure to live vocals and certainly differ greatly from the pure sine waves that were electronically created. Including an intermediary comparison group of autotuned vocals, which are composed mostly of pure sine waves but retain some timbre elements of vocals, was considered, but in the interest of simplicity, only two groups were used.

Considering these limitations, the implications of findings are narrow. A much larger sample size and more comprehensive training program would increase rigor and allow for a better comparison between memory trainings using vocal pitches and electronic pitches. In future experiments, it would be interesting to see if using multiple training sessions over a longer period of time would have more success in training performance on tonal memory tasks. It could also be useful to compare baseline tonal memory assessments across a larger variety of auditory stimuli, such as wind instruments, percussive pitches, and animal calls. In this case, no training sessions would be necessary to investigate effects of the timbre of a tone on its memorability. Further experimentation is still required to probe the link between vocally produced pitches and absolute pitch memory.

APPENDIX A: INFORMED CONSENT FORM

Appendices A, B, and C contain visual samples of the informed consent form, survey questions, and the pre-test explore page. Please note that access to complete experimental data is available by request at graysonfletcher29@gmail.com.



This experiment is being conducted by Grayson Fletcher as part of his Honors Thesis at the University of South Carolina, under the supervision of Dr. Jessica Green. If you have any questions about the experiment you can contact Grayson (fletcheg@email.sc.edu) or Dr. Green (greenjj3@mailbox.sc.edu). After you have completed the experiment you will be emailed a summary of the experiment, contact information for the researchers, and information about using the experiment for course extra credit.

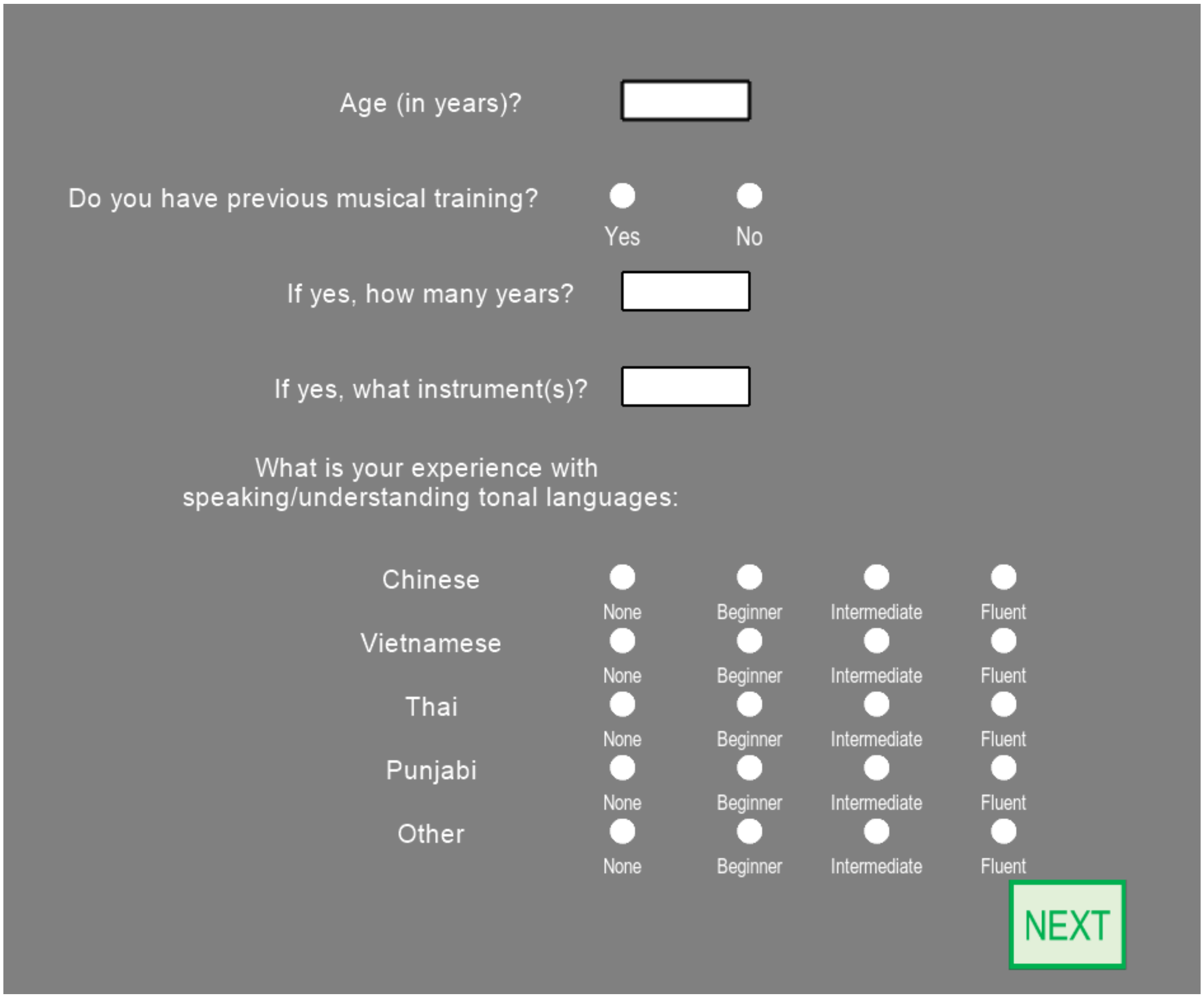
To continue to the experiment, please type your name and email address below and then click CONTINUE to begin. If at any time you wish to terminate your participation you can exit out of the experiment by pressing the escape key.

Name:

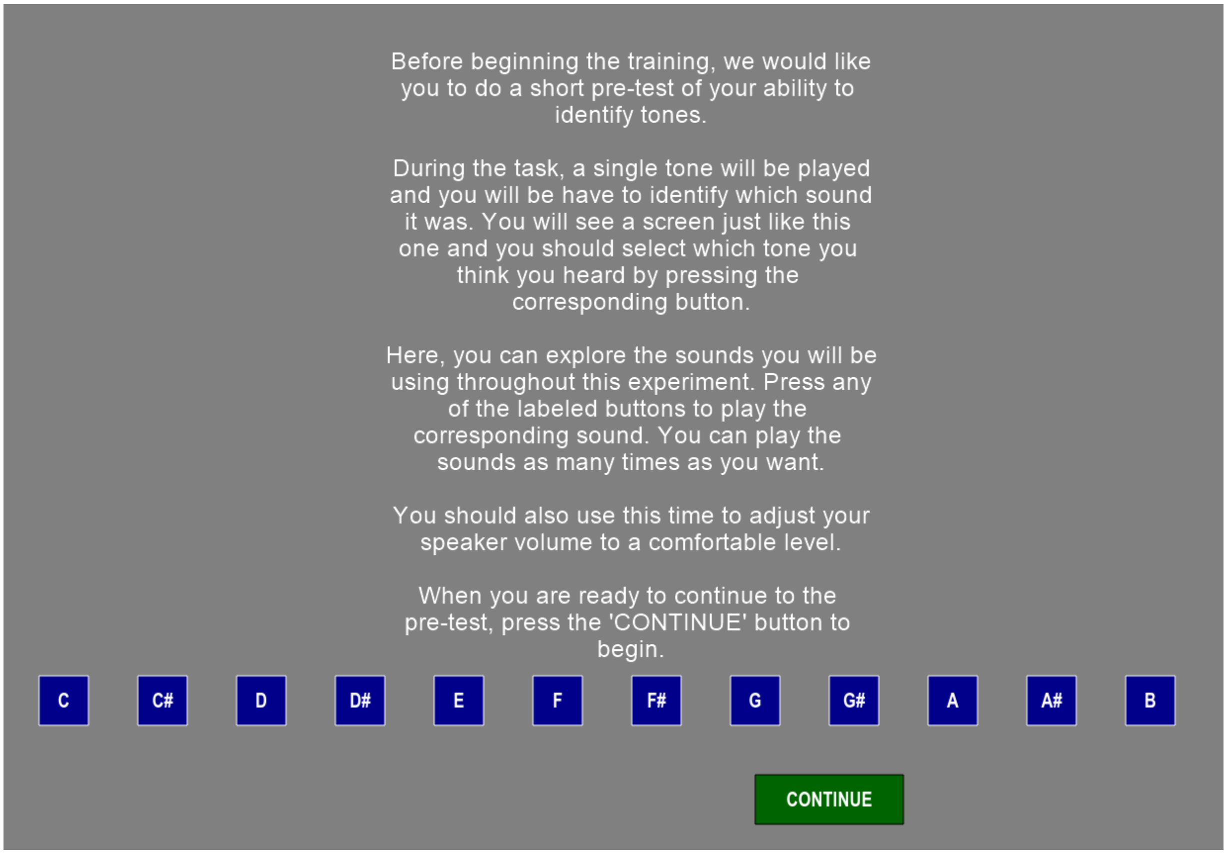
Email:

CONTINUE

APPENDIX B: DEMOGRAPHIC SURVEY



APPENDIX C: PRE-TEST EXPLORE PAGE



REFERENCES

Albouy, P., Schulze, K., Caclin, A., Tillmann, B. (2013). Does tonality boost short-term memory in congenital amusia? *Brain Research*, 1537, 224-232.

Bittrich, K., Schulze, K., Koelsch, S. (2012). Electrophysiological correlates of verbal and tonal working memory. *Brain Research*, 1432, 84-94.

Deutsch, D., Henthorn, T., & Dolson, M. (2004). Absolute pitch, speech, and tone language: Some experiments and a proposed framework. *Music Perception* 21, 339-356.

Eitan, Z., Ben-Haim, M.S., Margulis, E.H. (2017). Implicit absolute pitch representation affects basic tonal perception. *Music Perception*, 34 (5), 569–584.

Frey-Clark, M. (2017). Pitch systems and Curwen hand signs: A review of literature. Update: Applications of Research in Music Education, 36(1), 59-65.

Howland, Catherine E., "An Instrumental Approach to Sight Singing: The Use of Mock-Fingerings" (2020). *Senior Theses*. 562.

Moulton C. (2014). Perfect pitch reconsidered. *Clinical medicine (London, England)*, *14*(5), 517–519.

Schulze, K., Tillmann, B. (2013). Working memory for pitch, timbre, and words. *Memory*, 21(3), 377-395.

Takeuchi, A. H., & Hulse, S. H. (1993). Absolute pitch. *Psychological bulletin*, *113*(2), 345–361.

Wong, Y. K., Lui, K. F. H., Yip, K. H. M., & Wong, A. C. -N. (2020). Is it impossible to acquire absolute pitch in adulthood? *Attention, Perception, & Psychophysics, 82*(3), 1407–1430.

Wong, Y. K., Ngan, V. S., Cheung, L. Y., & Wong, A. C.-N. (2020). Absolute pitch learning in adults speaking non-tonal languages. *Quarterly Journal of Experimental Psychology*, *73*(11), 1908–1920.