# AN INVESTIGATION OF THE EFFECTS OF SUGGESTION IN AUDITORY MEMORY: HOW MUCH DO HUMANS REALLY LIKE THE SOUND OF OUR OWN VOICE?

By

Ethan Axler

Department of Psychology

Submitted in Partial Fulfillment of the requirements for the degree of Bachelor of Arts in Honours Psychology

Faculty of Arts and Social Science

Huron University College

London, Canada

April 30, 2024

HURON UNIVERSITY COLLEGE

**CERTIFICATE OF EXAMINATION**

Advisor: Dr. Stephen Van Hedger

Reader: Dr. Melissa Meade

The thesis by:

Ethan Axler

entitled:

An Investigation of the Effects of Suggestion in Auditory Memory: How Much Do Humans Really Like the Sound of our Voice?

is accepted in partial fulfilment of the requirements for the degree of Bachelor of Arts in Honours Psychology

| | |
|---|---|
| April 30th, 2024 | Dr. Irene Cheung |
| Date | Chair of Departmen |

**Abstract**

This study builds upon Weiss et al.'s (2012) findings that vocally produced melodies are better remembered than instrumental versions of the same melodies, an effect the authors referred to as the *vocal memory advantage*. Weiss et al. (2012) suggested that the *vocal memory advantage* might reflect preferential processing for biologically meaningful sounds. The present study measures whether listener expectations (i.e., expecting to hear a vocal versus instrumental sound) are sufficient to produce the *vocal memory advantage*. All participants listened to identical melodies produced by an ambiguous sound source. Participants in the vocal condition were told that these sounds were human whistles, whereas participants in the non-vocal condition were told that these were produced by wind instruments (e.g., a slide whistle). Participants heard half of the melodies in an encoding phase and then rated their confidence as to whether melodies were old or new in a recognition phase, during which old melodies were interspersed with new (“foil”) melodies. Despite not finding any evidence for the *vocal memory advantage*, analyses indicated both directly and indirectly that the manipulation was successful (i.e., participants believed the sounds originated from human whistles or from wind instruments). Although the study did not suggest that the *vocal memory advantage* can be engendered from listener expectations alone, the findings do support how instructional manipulations can influence other aspects of perception, emphasizing the delicate interplay between sensation and perception.

*Keywords*: Audition, Vocal, Instrument, Memory

**Acknowledgements**

I would like to express my sincere gratitude to my supervisor, Dr. Van Hedger, whose guidance and support were invaluable throughout the process of conducting and writing my thesis manuscript. Dr. Van Hedger's experience and expertise in the field transformed my research aspirations into a reality. His unwavering dedication and patience empowered me to pursue my interests with confidence. Working alongside Dr. Van Hedger has been a profound learning experience that will undoubtedly shape my academic journey. It was a true privilege to have him as a supervisor and mentor in Huron's Psychology Department.

I also extend my heartfelt appreciation to my second reader, Dr. Meade. Thank you for offering invaluable insights and constructive feedback on my research. Your support and thought-provoking questions played a pivotal role in refining my thesis into its current form.

Finally, to my family and friends – thank you for your unwavering support throughout my academic journey. Your encouragement and guidance have been instrumental in my success, including the completion of this thesis. I am immensely grateful for the positive input.

## Table of Contents

## Introduction

Music is a fundamental human activity, found in every known culture (Killin, 2018). Humans have been making musical instruments for at least 40,000 years, as evidenced by the discovery of bone flutes (Conard et al., 2009). Although the origin of vocally produced music is harder to verify archaeologically, the voice is arguably the most important musical instrument both historically and culturally (La Barbara, 2002), with several evolutionary theories suggesting that a music-like system of vocal communication (e.g., composed of prosodic sighs, grunts, and cries) was the foundation for modern systems of speech and music (Greenhill et al., 2017). The present study assesses whether musical signals perceived as originating from the voice show preferential processing, including enhanced memory.

Given the ubiquity and importance of music in contemporary and prehistoric times, it is perhaps not surprising that music has been used by researchers as a real-world "tool" for assessing fundamental aspects of human perception and cognition. Music has been used to assess basic principles of auditory attention (Strait et al., 2015), learning (Brown & Palmer, 2012), categorization (Marin et al., 2012), and, most importantly for the present work, memory (Cohen et al., 2011). When considering the specific relationship between music and memory, it is important to consider that both music and memory are multidimensional constructs. Music varies along several dimensions (e.g., instrumentation, tempo, tonality, genre), which leaves open the question of how these features of music influence memory processes. Similarly, memory is a multidimensional construct, operating along multiple timescales (from milliseconds to decades) and in both implicit and explicit domains. As such, it is important to consider the specific features of music and the specific memory construct being considered in appropriately characterizing the existing literature on music and memory.

One of the most frequently studied topics within the domain of music and memory is the relationship between music and autobiographical memory. Music can serve as a powerful retrieval cue for autobiographical memory, bringing one back to the sights and sounds of an event from across the lifespan (Jakubowski & Ghosh, 2021). The retrieval of autobiographical memory through musical cues may be facilitated by both the frequency with which people listen to music as well as the tradition to coupling culturally significant moments and events to music (Greasley & Lamont, 2011; North et al., 2004; Sloboda et al., 2001).

Cultural congruence has as well been shown to have a significant influence on musical memory. Kim and Jang (2016) demonstrated that individuals who had a cultural attachment to a piece of music were more successful than those who did not have a cultural attachment to a piece of music in their ability to recall a specific event or situation attached to the music. It should be noted that similar to the autobiographical memory literature, research pertaining to the cultural congruence of music has been investigated using associative retrieval cues. Music with a national (cultural) style of sound can evoke images and concepts congruent with the culture (North et al., 2016). For example, playing German music to a listener may evoke the listener to think of German stereotypes such as beer and bratwurst whereas a French piece of music may evoke an individual to think of champagne and baguettes (North et al., 2016).

Beyond this previous literature showing that music serves as a strong retrieval cue for associated memories, the features of music also tend to be remembered very well (Jancke, 2008). It has long been understood that experiences that elicit arousal are more likely to be remembered then experiences that do not elicit emotional arousal (Kensinger, 2009). Moreover, this emotional memory enhancement has been shown to be especially pronounced for music that elicits a strong sense of arousal (Kensinger, 2009). Therefore, the emotional salience of a piece of music plays an important role in emotionally arousing the listener further encouraging subsequent memory of the piece of music. In a study conducted by Särkämö et al. (2008), participants who were stroke victims who listened to their favorite music on a daily basis showed increased verbal memory and focused attention compared to participants who listened to audio books or no listening material at all. The study demonstrates that music can facilitate subsequent cognitive performance if it is aroused via the emotional salience of music. Furthermore, arousal from music brought on by emotional salience allows for not only focal memory improvements but subsequent cognitive performance. Music acts as a great delivery method of emotional information for its deep emotional salience among listeners (Alger et al., 2019). Emotional salience is a memory cue that allows for memory retrieval via prioritizing of information with emotional significance (Alger et al., 2019). Emotions have been shown to communicate recognition understanding processes in the brain so that the corresponding situations or objects can receive preferential attention allowing for extra processing resources in the brain (Perlovsky, 2012). Therefore, it is evident that emotional salience plays an integral role in memory retrieval in humans which is as well true for music. Emotional salience, however, is only one musical feature that influences memory of music.

Another musical feature to be shown to influence memory of music in general terms is the presence of lyrics (audible words) in the piece of music. Previous research has shown that the presence of lyrics in music, relative to music that does not contain lyrics, hinders performance on concurrent tasks presented alongside the music, including measures concurrently presented materials of verbal memory, visual memory and reading comprehension (Souza & Barbosa, 2023).

The notion that lyrics impair memory on concurrent tasks (memory tasks) could have multiple explanations. First, it is possible that in processing the linguistic meaning of lyrics, attention is drawn away from the concurrent memory tasks (Shtyrov et al., 2013), reflecting the notion that cognitive resources are finite (Ward et al., 2017). Therefore, the presence of lyrics in a piece of music may be a sufficiently salient signal to draw attention away from a concurrent task, resulting in poorer memory performance. This then suggests that we process music with lyrics in a differ manner to music without lyrics (Souza & Barbosa, 2023), possibly indicating an attentional bias toward vocal music. However, it is unclear whether this bias is explained by the linguistic content within lyrics, or whether it represents something more basic about processing the characteristic sound (i.e., the *timbre*) of the human voice.

There are several reasons to expect a more general attentional bias toward the human voice in music, regardless of whether it is producing understandable linguistic content. Within the non-human animal literature, there is ample evidence that species have preferential processing of same-species (conspecific) vocalizations (e.g., Fan et al., 2019). Under this view, what really matters is the source of the signal (i.e., whether it is a voice or not) - not necessarily whether there is a clear linguistic signal present (i.e., lyrics).

This concept was explored by Weiss et al. (2012) who tested the idea that the human voice might be preferentially processed and remembered better than non-vocal signals, even in situations where the voice is not producing a linguistic message (e.g., through lyrics). Weiss et al. (2012) presented participants with previously unfamiliar folk melodies across different instruments (within-participant). Some instruments were classified by the researchers as more familiar (voice and piano), and others were classified as less familiar (banjo and marimba). Participants were subsequently tested on previously heard melodies which were randomly mixed with novel melodies in a recognition test. The study demonstrated that participants better remembered vocal melodies compared to all other instruments.

The *vocal memory advantage* seen in Weiss et al. (2012) can potentially be understood using levels of processing theory (Craik & Lockhart, 1972). Levels of processing theory suggests that items that are processed more deeply (e.g., attended to and integrated into a richer conceptual network) will be better remembered than items that are processed more superficially. According to Weiss et al. (2012), levels of processing theory may underlie the vocal memory advantage, as vocal signals may increase listeners' arousal or vigilance due to their biological significance, thereby resulting in a greater depth of processing and enhanced memory. Weiss et al. (2012) found a vocal memory advantage without audible lyrics as the singer in the study sang the syllable "la" so their findings cannot be explained with the notion that the memory advantage was due to increased richness of the experience via the simultaneous presentation of a melody and lyrics. The lack of the extra channel of information (language) supports the idea that there is a biological advantage to processing the timbre of the human voice. A complementary perspective is that listeners have considerable expertise listening to voices, akin to the expertise in processing faces (e.g., Kanwisher et al., 1997). The human voice is without a doubt more recognizable and familiar than that of instrumental timbres (e.g., piano). When one listens to the human voice in music, listeners may infer personal characteristics from this signal (e.g., age, sex) that would not be available with a non-vocal instrument, which could also facilitate memory via levels of processing (Craik & Lockhart, 1972).

Within this framework that the human voice may bias attention and thus influence musical memory, there is an open question as to whether listener *expectations* (i.e., expecting to hear a vocal or non-vocal signal) are sufficient to observe memory advantages. This is an important theoretical question to address, as it helps answer whether the *vocal memory advantage* is better explained by inherent acoustic differences between vocal and non-vocal signals, which may bias attention, or whether these biases in attention can be driven by the mere expectation of hearing a vocal signal (i.e., top-down effects). In other auditory domains, listener expectations have been shown to radically alter perceptual processing of a given signal (Kafkas & Montaldi, 2018). Past literature has shown that processing of sine-wave speech - an ambiguous speech signal that can either be heard as speech or non-speech “whistles” - changes when it is heard as speech compared to when it is heard as non-speech (Tremblay et al., 2000). An additional piece of evidence that illustrates the importance of listener expectations is the “speech to song” illusion, in which the same speech signal - when repeated - is sometimes heard by listeners as being sung following this repetition (Deutsch et al., 2011). Tiernay et al. (2013) found that excerpts that transformed from speech to song activated more extensive brain networks than repeated speech that did not transform to song. These findings are of particular relevance for the present study, as they suggest that the same acoustic signal may be processed differently in a musical context when it is heard as being sung. In other words, the findings of Tierney et al. (2013) suggest that how a listener interprets the sound is possibly just as important as the perceptual features of the sound in predicting how it is processed, which could have implications for how the sounds are subsequently remembered (even though this was not directly tested by the authors).

The present study teases apart whether the vocal memory advantage is influenced by listener expectations (i.e., telling participants the signal originates from the voice), when holding the acoustic signal constant. Participants will be randomly assigned to one of two conditions (vocal condition or non-vocal condition). The only difference between the conditions will be the instructions the participants receive. Although all participants will hear the same melodies, the acoustic signal itself was designed to be ambiguous and able to be interpreted as either originating from human whistles (a vocal signal) or from wind instruments (a non-vocal signal). Thus, participants in the vocal condition will be provided with instructions that these melodies were whistled, whereas participants in the non-vocal condition will be provided with instructions that these melodies were played on wind instruments (e.g., a tin whistle). Given that the only difference between conditions is the manipulation of the instructions, the present study is well established to address whether participant expectations are sufficient to drive the *vocal memory advantage* (i.e., while holding the acoustics of the sound constant).

In line with previous literature (e.g., Weiss et al., 2012; 2017), it is hypothesized that participants who are in the vocal conditions will have better memory of the melodies than participants who are in the non-vocal condition. Such findings would suggest that participant expectations of hearing a vocally produced signal are sufficient to enhance memory for musical melodies.

# Method

## Participants

A total of 66 participants completed the study and 60 (*M* = 19.08 years old, *SD* = 3.47, range of 17 to 38 years old, 20 men, 39 women, 1 prefer not to answer) were retained for analysis. Upon accessing the study, participants were randomly assigned to either the vocal condition (*n* = 33) or the instrumental condition (*n* = 27). Participants were excluded from analyses if they failed to respond to more than two of four simple auditory attention checks (see *Procedure* for details), which were embedded in the encoding and recognition tasks. Participants were recruited primarily from introductory psychology courses with a research participation course component (Psychology 1100 and 1000) at Huron University College. We also allowed for “snowballing” as a potential recruitment strategy, in which recruited participants notified friends or acquaintances who meet the eligibility requirements about the study; however, any participants who were recruited outside of Psychology 1100 or 1000 did not receive compensation for completing the study. Participants all had self-reported normal hearing and vision. The only exclusion criteria of the study where participants must not have any neurological disorder (e.g., schizophrenia).

**Materials**

The study was programmed in jsPsych 7 (de Leeuw, 2015), and participants were able to access the study from their personal computers. The 24 melodies used in the memory task were selected by the author from a database of folk songs (https://www.8notes.com/digital_tradition/) from the United Kingdom. Folk songs were primarily selected based on presumed unfamiliarity and length, but also naturally contained variability in tempo and mode. The songs were approximately 25 seconds in length (range of 20 to 33 seconds), depending on the melody. A full list of the folk songs used in the study is reported in the Appendix. Selected folk melodies were then played on a digital keyboard by one of the researchers using the “Whistle 2.0” virtual instrument software (Rast Media GmbH: https://rastsound.com/). The parameters of the virtual instrument were adjusted by the researchers until both were in agreement that the resulting sound could be reasonably heard as either a human whistle or a wind instrument depending on context. Additionally, to corroborate the beliefs of the researchers, all participants were asked how realistic the sounds were for their respective conditions (e.g., how realistic the sounds were as whistles). The filler task presented participants with the Sandia Matrices (Matzen et al., 2010), which are a free alternative to the Raven's Matrices.

## Procedure

### *Auditory Calibration*

Participants first completed a short auditory calibration. This consisted of two components - a volume adjustment and a headphone assessment. The volume adjustment required participants to click on a button presented in the middle of the computer screen, which triggered the playing of a short musical excerpt, which was normalized to the same amplitude level as the melodies used in the study. Participants were instructed to adjust their computer's volume to a level where the short excerpt was at a comfortable listening level. Participants could replay the short excerpt if desired. Once participants were satisfied with the volume level, they pressed another button to continue.

Next, participants were presented with a short loudness judgment task as described by Woods et al. (2017), which was meant to assess headphone use. On each trial, participants would hear three tones, and had to determine which of the three tones was quietest. This assessment was designed to be easy for those using headphones, and difficult for those listening to the tones without the use of headphones. There were six total trials and responding correctly on at least five trials was taken as evidence of headphone use as recommended by Woods et al. (2017). Of the 53 included participants in the analyses, 24 passed the headphone assessment. However, the use of headphones was encouraged but not required, and therefore headphone assessment performance was not used as an exclusion criteria.

***Instructional Manipulation***

Following auditory calibration, participants were introduced to the main task of listening to unfamiliar folk melodies. The general flow of the instructions was identical for participants in the vocal and instrumental conditions; however, the specific framing of the instructions and the presented instructional video varied across conditions. Specifically, participants were either told that they would hear *whistled* melodies (vocal condition), which was accompanied by a synchronized video of the two researchers whistling a five-note melody or told that they would hear *played* melodies (instrument condition), which was accompanied by a synchronized video of the two researchers playing a five-note melody on wind instruments (a tin whistle and a slide whistle). Both videos showed the two researchers simultaneously via a split screen. Importantly, the videos were edited such that the naturally produced audio from whistling and playing the wind instruments was replaced with the same ambiguous timbre used for the main memory assessment. In this sense, participants across both conditions heard the same thing during the instructions; the only difference was in how the sounds were framed, which occurred in the text and was further reinforced by the visuals of the video. Participants watched the video for their respective condition twice before moving onto the main melody encoding task.

***Melody Encoding Task***

Participants listened to 12 of the total 24 melodies (randomly selected) during the initial encoding phase. Each melody was presented twice. Following the second presentation of the melody, participants were asked to identify the perceived emotional valence of the melody (using a forced-choice response of *Sad, Neutral,* or *Happy*), which was not of primary interest in the present study but was implemented to keep participants engaged in listening to the melodies. At two points during the encoding phase (once within the first six melodies, once within the final six melodies) participants were given an auditory attention check. Attention checks consisted of short auditory prompts that asked the participants to press a key on their keyboard within an allotted timeframe (10 seconds), demonstrating they were paying attention. If participants did not press the designated key within the timeframe, the study moved on automatically.

### *Filler Task*

Following melody encoding, participants completed a short (5-minute) visual matrix reasoning task (Matzen et al., 2010). Participants were given up to 42 matrix reasoning problems, in which they saw a geometric pattern arranged in a 3 x 3 grid. The lower right portion of the grid was missing, and participants had to select the pattern that completed the matrix by clicking on one of eight images presented on the screen. After the allotted duration of five minutes, the script automatically moved on to the melody recognition task. The matrix reasoning task primarily served as a filler task between melody encoding and recognition, and thus performance is not considered in the analyses of the present study.

### *Melody Recognition Task*

Following the matrix reasoning filler task, participants were introduced to the melody recognition task. In the recognition task, participants listened to all 24 melodies - 12 new (“foil”) melodies that had not been presented during encoding, and 12 old (“target”) melodies that had been previously presented during encoding. Melody order was randomized. Following each melody, participants were asked to indicate their confidence as to whether the melody was new or old on a Likert scale, ranging from 1 (*Definitely new*) to 7 (*Definitely old*). This approach of measuring recognition memory using a confidence rating was identical to the approach used in the original vocal memory advantage work by Weiss et al. (2012). Participants were additionally asked to indicate how much they liked the melody on a Likert scale ranging from 1 (*Not at all*) to

7 (*Extremely*). Similar to the melody encoding phase, participants were presented with two auditory attention checks (one during the first 12 trials, one during the final 12 trials) during the melody recognition phase.

***Questionnaire***

Following the melody recognition phase, participants provided basic demographic information (age and gender, and highest level of completed education). Participants were additionally asked if they had received any training on a musical instrument (including the voice). If participants answered *yes*, they were asked to provide the age at which they began musical training, their primary musical instrument, how many years they had played their primary instrument, and how often they currently play music (answered categorically, with response options of *I am no longer musically active, Less than 0.5 hours/wk, 0.5-1 hour/wk, 1-3 hours/wk*, *3-5 hours/wk*, and *5+ hours/wk*). All participants, regardless of whether they reported musical training, were asked to indicate whether they had absolute pitch and whether they considered English to be their native language (and if not, to indicate their Native language as a free response).

Following these demographics, musical, and language questions, participants were asked how realistic the melodies sounded as either human whistles (vocal condition) or wind instruments (instrument condition). Participants made these responses on a Likert scale ranging from 1 (*Not at all*) to 5 (*Extremely*). Lastly, participants were asked in two sentences to describe what they thought the purpose of the study was. Participants were then provided with a debriefing letter, which they could download, and were then redirected to a digital credit receipt where they could enter personal information (not tied to the study data) to receive course credit.

### Data Analysis

The primary memory measure was calculated by subtracting participants' confidence ratings for new (foil) melodies from their confidence ratings for old (target) melodies, with a higher value reflecting better memory for the initial melodies heard in encoding (i.e., higher ratings for old versus new melodies). This memory score was first assessed against chance performance using a one-sample *t*-test against chance performance (i.e., a score of 0). An independent-samples *t*-test was then used to assess whether memory performance differed across conditions, as hypothesized.

Independent-samples *t*-tests were additionally used to assess whether participants in the vocal versus instrumental conditions differed (1) in how much they liked the melodies, (2) in the perceived quality of the sounds as representing either whistling or wind instruments, and (3) the perceived emotional valence of the melodies (taken from the ratings made during encoding). Given that the emotional valence ratings were made as a forced choice rating with three options (sad, neutral, happy), these values were recoded as -1, 0, and 1, respectively, which facilitated the calculation of a mean score for each participant. An independent samples *t*-test was also used to assess whether musicians differed from non-musicians in terms of memory performance. Finally, a paired-samples *t*-test was used to assess whether participants differed in their liking ratings for old versus new melodies.

## Results

### Memory Performance

Overall memory performance was significantly above chance, $t(59) = 6.75$, $p < .001$. Old melodies had a mean rating of 4.64 (*SD*: 0.66), whereas new melodies had a mean rating of 3.52 (*SD*: 0.87). Although participants in the vocal condition had nominally higher memory scores

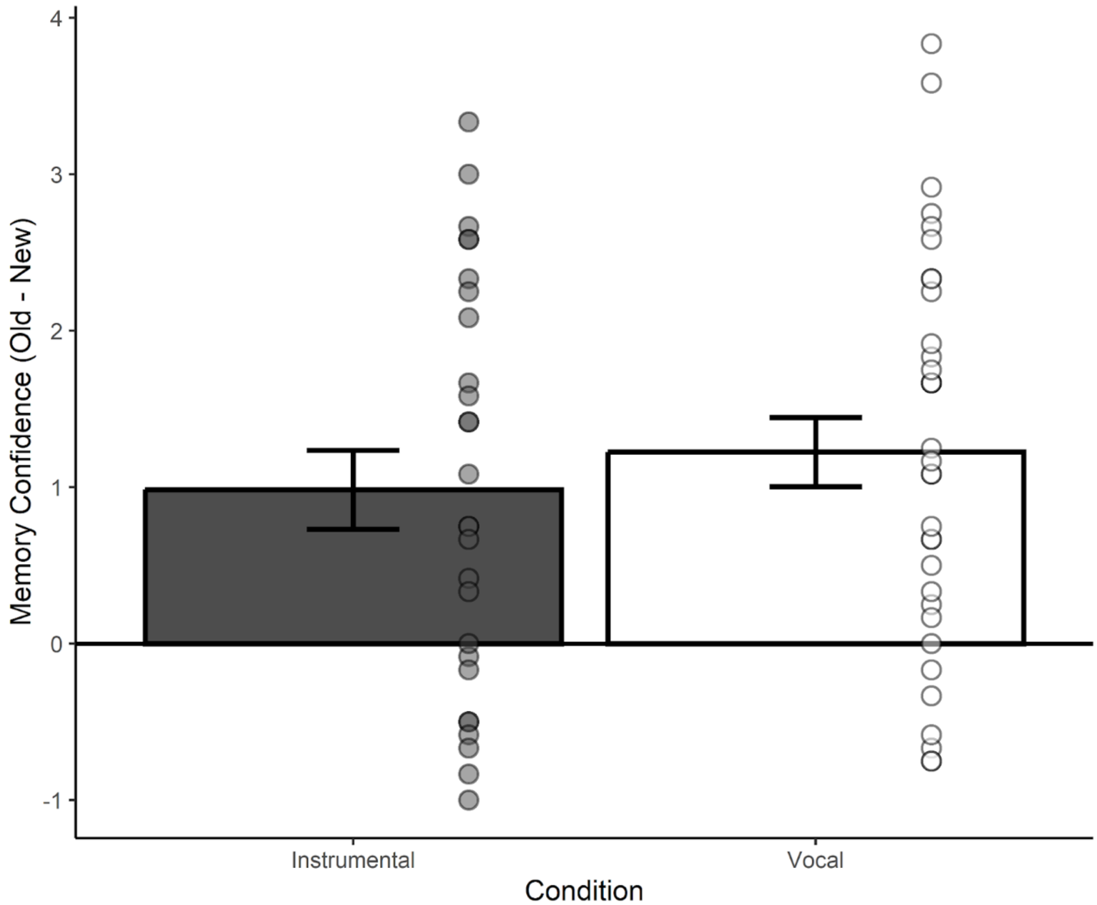
(*M*: 1.22, *SD*: 1.27) compared to participants in the instrumental condition (*M*: 0.98, *SD*: 1.31), this difference was not statistically significant, $t(58) = 0.72$, $p = .475$. Figure 1 plots memory performance as a function of condition. Memory performance did not differ between musicians (*M*=1.28, *SD*=1.30) and non-musicians (*M*=0.86, *SD*=1.24), $t(58) = 1.25$, $p = 216$.

**Liking Ratings**

Participants liked old melodies significantly more than new melodies, $t(59) = 3.71$, $p < .001$. The mean liking rating for old melodies was 4.58 (*SD*: 0.70) and the mean liking rating for new melodies was 4.27 (*SD*: 0.77). There was no significant difference in liking ratings as a function of condition, $t(58) = -0.06$, $p = .955$. The mean liking rating for participants in the vocal condition was 4.42 (*SD*: 0.57) and the mean liking rating for participants in the instrumental condition was 4.43 (*SD*: 0.77). There was, however, a significant difference in the difference between liking ratings for old and new melodies as a function of condition, $t(58) = 2.16$, $p = .035$. This significant effect was characterized by a more pronounced “old-versus-new” preference for participants in the vocal condition compared to participants in the instrumental condition (see Figure 2).

**Figure 1**

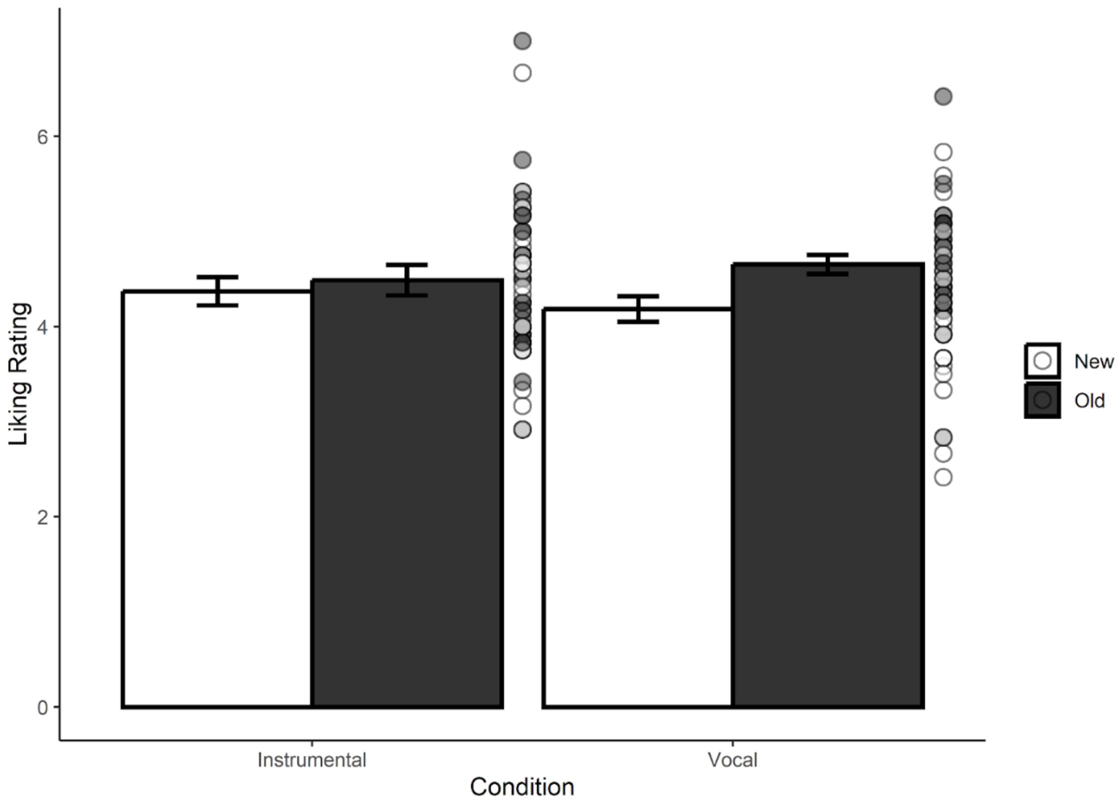
Memory scores for both vocal and instrumental conditions




*Note*: Error bars represent plus or minus one standard error of the mean. Circles represent individual participant scores.

**Figure 2**

Liking ratings as a function of melody type (old, new) and condition (vocal, instrumental)




*Note*: Error bars represent plus or minus one standard error of the mean. Circles represent individual participant scores.

**Perceived Quality**

There were no significant differences in perceived quality of the melodies in depicting their respective categories (whistling for the vocal condition, wind instruments in the instrumental condition), $t(58) = 0.15$, $p = .882$. The mean quality rating for participants in the vocal condition was 3.48 (*SD*: 0.94), and the mean quality rating for participants in the instrumental condition was 3.44 (*SD*: 1.15).

**Emotion Ratings**

Participants in the vocal condition rated the melodies presented in encoding significantly lower than participants in the instrument condition, $t(58) = -2.14$, $p = .037$. Given the calculation of the emotion rating score, this means that participants in the vocal condition judged the encoding melodies are overall sadder than participants in the instrument condition. The mean rating for participants in the vocal condition was 1.11 (*SD*: 0.22), and the mean rating for participants in the instrumental condition was 1.24 (*SD*: 0.27).

## Discussion

The current study did not replicate the *vocal memory advantage* reported in Weiss et al. (2012) There were no significant memory differences as a function of condition, despite some participants being told that the sounds were produced by the human voice. One possible reason for why the Weiss et al (2012) study was not able to be replicated in the current study is because participants simply did not find the instructional manipulation convincing. Although it is very well possible that participants simply did not believe the manipulation (suggesting different origins of the sounds), there are reasons to believe that participants actually did find the manipulation convincing. When participants were asked to rate the quality of the sounds as representing either whistling or wind instruments, participants in both groups gave nearly identical ratings (3.48 in the vocal condition and 3.44 in the instrumental condition), suggesting that the participants in both conditions believed the origins of the sounds to a comparable degree. Moreover, it should be noted that the absolute ratings were fairly high, with mean ratings above the midpoint of the scale (as quality ratings were made on a 5-point scale) Taken together, this suggests that participants overall were reasonably convinced that the sounds they were hearing were either a wind instrument or a vocal whistle. Hence, the notion that the *vocal memory advantage* was absent because participants did not believe the ambiguous sounds were produced via the human voice (vocal condition) is not a sound explanation for the lack of replication of the Weiss et al. (2012) findings.

Beyond the evidence that people believed the manipulation in the current study, the memory scores were significantly above zero, indicating that people remembered the old versus new melodies in the study. This is another encouraging finding in the data, as this demonstrates that participants understood the task (despite being conducted online) and remembered novel folk melodies despite only hearing them twice in encoding. It should be noted that the melodies selected for the current study were different to those used in Weiss et al. (2012), which opened the possibility that the present folk melodies were too long, too complicated or too similar sounding to be appropriate for the present memory paradigm. However, the finding that participants were above chance in memory recognition suggests that the lack of the *vocal memory advantage* in the present study is not likely due to the selected folk melodies.

Unlike the Weiss et al. (2012) findings, the current study found no evidence that the vocal condition melodies were liked less. However, similar to Weiss et al. (2012) and many other findings in the memory literature, we found a familiarity effect on liking ratings. This means that melodies that were heard in encoding were liked more than the “foil” melodies that were heard for the first time during the recognition task. These findings support the mere-exposure effect found in the memory literature (e.g., Zajonc, 1980). The mere-exposure effect refers to the well-established finding that people evaluate a stimulus more positively after repeated exposure to the stimulus, even in the absence of explicit awareness that the stimulus has been previously encountered (Van Dessel et al., 2017). In the present study, participants may have liked the songs they heard previously because these melodies were easier to process, in line with theories of perceptual fluency (e.g, McKean et al., 2020). Perceptual fluency refers to the ease and speed with which an individual processes sensory information or stimuli (Reber et al., 1998). Therefore, when a stimulus is easily processed people tend to find the stimulus more positively likable (Reber et al., 1998). Perceptual fluency further explains why participants in the current study liked the melodies that were heard in encoding more than the melodies heard for the first time during the recognition task.

An unexpected finding in the current study was that the “familiarity effect” on liking was stronger in the vocal condition compared to that of the instrument condition. This finding may further support the idea that the participants genuinely heard these whistle sounds as sounds produced by a human in the vocal condition. A study by Souza et al. (2013) suggests when a speech is spoken by a familiar talker (e.g., a spouse or close friend), it is better understood compared to when the same content is spoken by a stranger. Applying this understanding to the current study, participants may have liked old melodies in the vocal condition more than those in the instrument condition as participants in the vocal condition may have been attending more to the personal characteristics of the whistler, whereas participants in the instrumental condition may have been attending to the qualities of the instruments themselves (rather than who was playing the instruments). Put another way, given that the instrument in the vocal condition was the human voice, this may have encouraged the participant to attend to personal characteristics (e.g., is it from that of a young male or an older male voice?). In regard to the current study, even though all the participants saw similar videos (of the same people either whistling or playing wind instruments), participants in the vocal condition may draw more on personal characteristics (e.g., remembering who is whistling), whereas participants in the instrumental condition might be focusing their attention more on the instruments themselves. Although, it should be noted this evidence is speculative, this is one possible reason why participants in the vocal condition might like old melodies more than participants in the instrumental condition.

Another unexpected finding, which suggests that participants believed the instructional manipulation, was the significant difference in perceived emotion ratings across conditions found in encoding. Specifically, participants in the vocal condition rated melodies as overall sadder than participants in the instrumental condition. The musicology literature demonstrates that voices are typically rated as being able to convey sadness much more effectively than (high pitched) wind instruments. For instance, Huron et al. (2014) showed that the voice has a higher capacity for conveying sadness with a mean sadness capacity of 6.91 (*SD*: 0.44) while the piccolo (the highest pitched wind instrument in an orchestra, and the closest analogue to the wind instruments used in the instructional manipulation in the current study) has a mean sadness capacity of 3.33 (*SD*: 1.74). Although the measure of a melody's ability to convey sadness is not the focus of the current study, this finding once again indirectly suggests that the manipulation in the current study was successful.

Given this converging evidence that the manipulation was successful, it is perhaps more puzzling that there was no observed *memory vocal advantage*. One explanation for these findings is that whistles are not a good means of conveying rich personal information. In the initial description of the *vocal memory advantage*, Weiss et al. (2012) discussed how voices may afford a deeper level of processing than non-vocal instrumental sounds. However, whistles are not produced through vibrations of the vocal cords and generally do not contain rich harmonic structures. This means that personal characteristics (e.g., age, sex) are not readily apparent from whistles (unlike sounds produced via the vocal cords). Therefore, whistles may not be able to capture the richness of the biological signals that sufficiently capture attention and enhance memory performance through affording a deeper level of processing.

A future direction for the current study would be to compare the whistling sounds to sung versions of the same folk melodies. This would firstly allow for a more direct replication of Weiss et al. (2012) with a different corpus of melodies and would additionally allow comparisons of two human signals (singing and whistling) in terms of memory performance. This approach would also allow the testing of the possibility that whistles (compared to singing) do not afford the same inferences of personal characteristics, as researchers could ask participants to make judgments on the producer's age, sex, etc. for both the sung and whistled sounds. The prediction is that participants would be more consistent in their judgments of personal characteristics for the sung melodies compared to the whistled melodies, further supporting the conjecture that the lack of a *vocal memory advantage* in the present study may be due to whistles being unsuited to inferring personal characteristics. There are some limitations of the current study that could be addressed in future studies**.** The online nature of the study may have introduced too much variance (e.g., in listening environments) to be able to observe a *vocal memory advantage*. As such, future iterations of the study may consider enacting in-person testing for data collection. This would allow for an additional level of control for the study that would lessen the chance of a type II error. With this said, the removal of participants who failed the auditory attention checks in the current study provides some assurance that the analyzed participants were engaged with the task.

In sum, the present study investigated whether instructions of telling participants they were hearing a vocal signal (compared to a non-vocal signal) was sufficient to elicit a *vocal memory advantage*, despite the auditory signals being identical across conditions. The current study found no evidence for a *vocal memory advantage* as a function of instructional manipulation; however, the current study demonstrated that the manipulation used in the study was successful and influenced several aspects of responses, including the strength of the familiarity-liking advantage and even the emotional categorization of the melodies**.** Although the study yielded a null hypothesis, the study has important implications and has offered clear directions for future research. These implications include the notion that participants can be pushed via instructional manipulation to believe different origins of a sound, even if in reality the sounds are identical. This demonstrates the malleability in musical perception in regard to a melody’s origin and may have resulting changes in perception and decision making once a participant has decided what they think the origin of a sound is. Although the study did not support the idea that instructional manipulations can influence auditory memory, a sound's origin nevertheless has an effect on one's perception of said sound which further demonstrates the important balance between sensation and perception.

**References**

Alger, S. E., Chen, S., & Payne, J. D. (2019). Do Different Salience Cues Compete for Dominance in Memory over a Daytime Nap? *Neurobiology of Learning and Memory*, *160*, 48–57. https://doi.org/10.1016/j.nlm.2018.06.005

Brown, R. M., & Palmer, C. (2012). Auditory–motor learning influences auditory memory for music. *Memory & Cognition*, *40*(4), 567–578. https://doi.org/10.3758/s13421-011-0177-x

Cheng, F.-Y., Xu, C., Gold, L., & Smith, S. (2021). Rapid Enhancement of Subcortical Neural Responses to Sine-Wave Speech. *Frontiers in Neuroscience*, *15*. https://www.frontiersin.org/articles/10.3389/fnins.2021.747303

Cohen, M. A., Evans, K. K., Horowitz, T. S., & Wolfe, J. M. (2011). Auditory and visual memory in musicians and nonmusicians. *Psychonomic Bulletin & Review*, *18*(3), 586–591. https://doi.org/10.3758/s13423-011-0074-0

Conard, N. J., Malina, M., & Münzel, S. C. (2009). New flutes document the earliest musical tradition in southwestern Germany. *Nature*, *460*(7256), Article 7256. https://doi.org/10.1038/nature08169

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, *11*(6), 671–684. https://doi.org/10.1016/S0022-5371(72)80001-X

de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods*, *47*(1), 1–12. https://doi.org/10.3758/s13428-014-0458-y

*Evolutionary dynamics of language systems*. (n.d.). https://doi.org/10.1073/pnas.1700388114

Fan, Y., Yue, X., Yang, J., Shen, J., Shen, D., Tang, Y., & Fang, G. (2019). Preference of spectral features in auditory processing for advertisement calls in the music frogs. *Frontiers in Zoology*, *16*(1), 13. https://doi.org/10.1186/s12983-019-0314-0

*Functions of Music in Everyday Life: An Exploratory Study Using the Experience Sampling Method—John A. Sloboda, Susan A. O'Neill, Antonia Ivaldi, 2001*. (n.d.). Retrieved November 25, 2023, from https://journals.sagepub.com/doi/10.1177/102986490100500102

Greasley, A. E., & Lamont, A. (2011). Exploring engagement with music in everyday life using experience sampling methodology. *Musicae Scientiae*, *15*(1), 45–71. https://doi.org/10.1177/1029864910393417

Greenhill, S. J., Wu, C.-H., Hua, X., Dunn, M., Levinson, S. C., & Gray, R. D. (2017). Evolutionary dynamics of language systems. *Proceedings of the National Academy of Sciences*, *114*(42), E8822–E8829. https://doi.org/10.1073/pnas.1700388114

Huron, D., Anderson, N., & Shanahan, D. (2014). "You Can't Play a Sad Song on the Banjo:" Acoustic Factors in the Judgment of Instrument Capacity to Convey Sadness. *Empirical Musicology Review*, *9*(1), Article 1. https://doi.org/10.18061/emr.v9i1.4085

Jakubowski, K., & Ghosh, A. (2021). Music-evoked autobiographical memories in everyday life. *Psychology of Music*, *49*(3), 649–666. https://doi.org/10.1177/0305735619888803

Jäncke, L. (2008). Music, memory and emotion. *Journal of Biology*, *7*(6), 21. https://doi.org/10.1186/jbiol82

Kafkas, A., & Montaldi, D. (2018). Expectation affects learning and modulates memory experience at retrieval. *Cognition*, *180*, 123–134. https://doi.org/10.1016/j.cognition.2018.07.010

Kensinger, E. A. (2009). Remembering the Details: Effects of Emotion. *Emotion Review*, *1*(2), 99–113. https://doi.org/10.1177/1754073908100432

Killin, A. (2018). The origins of music: Evidence, theory, and prospects. *Music & Science*, *1*, 2059204317751971. https://doi.org/10.1177/2059204317751971

Kim, J.-H., & Jang, S. (Shawn). (2016). Memory Retrieval of Cultural Event Experiences: Examining Internal and External Influences. *Journal of Travel Research*, *55*(3), 322–339. https://doi.org/10.1177/0047287514553058

La Barbara, J. (2002). Voice is the original instrument. *Contemporary Music Review*, *21*(1), 35–48. https://doi.org/10.1080/07494460216648

Marin, M. M., Gingras, B., & Stewart, L. (2012). Perception of musical timbre in congenital amusia: Categorization, discrimination and short-term memory. *Neuropsychologia*, *50*(3), 367–378. https://doi.org/10.1016/j.neuropsychologia.2011.12.006

Matzen, L. E., Benz, Z. O., Dixon, K. R., Posey, J., Kroger, J. K., & Speed, A. E. (2010a). Recreating Raven's: Software for systematically generating large numbers of Raven-like matrix problems with normed properties. *Behavior Research Methods*, *42*(2), 525–541. https://doi.org/10.3758/BRM.42.2.525

Matzen, L. E., Benz, Z. O., Dixon, K. R., Posey, J., Kroger, J. K., & Speed, A. E. (2010b). Recreating Raven's: Software for systematically generating large numbers of Raven-like matrix problems with normed properties. *Behavior Research Methods*, *42*(2), 525–541. https://doi.org/10.3758/BRM.42.2.525

Mcadamsand, S., & Bigand, E. (1993). Introduction to auditory cognition. In S. McAdams & E. Bigand (Eds.), *Thinking in Sound: The Cognitive Psychology of Human Audition* (p. 0). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198522577.003.0001

McKean, B., Flavell, J. C., Over, H., & Tipper, S. P. (2020). Three minutes to change preferences: Perceptual fluency and response inhibition. *Royal Society Open Science*, *7*(10), 200766. https://doi.org/10.1098/rsos.200766

North, A. C., Hargreaves, D. J., & Hargreaves, J. J. (2004). Uses of Music in Everyday Life. *Music Perception*, *22*(1), 41–77. https://doi.org/10.1525/mp.2004.22.1.41

North, A. C., Sheridan, L. P., & Areni, C. S. (2016). Music Congruity Effects on Product Memory, Perception, and Choice. *Journal of Retailing*, *92*(1), 83–95. https://doi.org/10.1016/j.jretai.2015.06.001

Perlovsky, L. (2012). Cognitive function, origin, and evolution of musical emotions. *Musicae Scientiae*, *16*(2), 185–199. https://doi.org/10.1177/1029864912448327

Reber, R., Winkielman, P., & Schwarz, N. (1998a). Effects of Perceptual Fluency on Affective Judgments. *Psychological Science*, *9*(1), 45–48.

Reber, R., Winkielman, P., & Schwarz, N. (1998b). Effects of Perceptual Fluency on Affective Judgments. *Psychological Science*, *9*(1), 45–48.

Scarinci, N., Worrall, L., & Hickson, L. (2008). The effect of hearing impairment in older people on the spouse. *International Journal of Audiology*, *47*(3), 141–151. https://doi.org/10.1080/14992020701689696

Shtyrov, Y., Goryainova, G., Tugin, S., Ossadtchi, A., & Shestakova, A. (2013). Automatic processing of unattended lexical information in visual oddball presentation: Neurophysiological evidence. *Frontiers in Human Neuroscience*, *7*. https://www.frontiersin.org/articles/10.3389/fnhum.2013.00421

Souza, A. S., & Barbosa, L. C. L. (2023). *Should We Turn off the Music? Music with Lyrics Interferes with Cognitive Tasks* (1). *6*(1), Article 1. https://doi.org/10.5334/joc.273

Souza, P., Gehani, N., Wright, R., & McCloy, D. (2013). The advantage of knowing the talker. *Journal of the American Academy of Audiology*, *24*(8), 689–700. https://doi.org/10.3766/jaaa.24.8.6

Stalinski, S. M., & Schellenberg, E. G. (2013). Listeners remember music they like. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(3), 700–716. https://doi.org/10.1037/a0029671

Strait, D. L., Slater, J., O'Connell, S., & Kraus, N. (2015). Music training relates to the development of neural mechanisms of selective auditory attention. *Developmental Cognitive Neuroscience*, *12*, 94–104. https://doi.org/10.1016/j.dcn.2015.01.001

Tierney, A., Patel, A. D., Jasmin, K., & Breen, M. (2021a). Individual differences in perception of the speech-to-song illusion are linked to musical aptitude but not musical training. *Journal of Experimental Psychology: Human Perception and Performance*, *47*(12), 1681–1697. https://doi.org/10.1037/xhp0000968

Tierney, A., Patel, A. D., Jasmin, K., & Breen, M. (2021b). Individual differences in perception of the speech-to-song illusion are linked to musical aptitude but not musical training. *Journal of*

*Experimental Psychology: Human Perception and Performance*, *47*(12), 1681–1697. https://doi.org/10.1037/xhp0000968

Tremblay, S., Nicholls, A. P., Alford, D., & Jones, D. M. (2000). The irrelevant sound effect: Does speech play a special role? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(6), 1750–1754. https://doi.org/10.1037/0278-7393.26.6.1750

Van Dessel, P., Mertens, G., Smith, C. T., & De Houwer, J. (2017). The mere exposure instruction effect: Mere exposure instructions influence liking. *Experimental Psychology*, *64*(5), 299–314. https://doi.org/10.1027/1618-3169/a000376

Ward, A. F., Duke, K., Gneezy, A., & Bos, M. W. (2017). Brain Drain: The Mere Presence of One's Own Smartphone Reduces Available Cognitive Capacity. *Journal of the Association for Consumer Research*, *2*(2), 140–154. https://doi.org/10.1086/691462

Weiss, M. W., Schellenberg, E. G., & Trehub, S. E. (2017). Generality of the Memory Advantage for Vocal Melodies. *Music Perception*, *34*(3), 313–318. https://doi.org/10.1525/mp.2017.34.3.313

Weiss, M. W., Trehub, S. E., & Schellenberg, E. G. (2012). Something in the Way She Sings: Enhanced Memory for Vocal Melodies. *Psychological Science*, *23*(10), 1074–1078. https://doi.org/10.1177/0956797612442552

Wood, K. C., Town, S. M., Atilgan, H., Jones, G. P., & Bizley, J. K. (2017). Acute Inactivation of Primary Auditory Cortex Causes a Sound Localisation Deficit in Ferrets. *PLOS ONE*, *12*(1), e0170264. https://doi.org/10.1371/journal.pone.0170264

# Appendix

List of folk songs used in the study.

| Song Name |
| --- |
| A Hundred Years Ago |
| Balaclava |
| Banks of the Dee (Parody) |
| Can Ye Sew Cushions? |
| Candlemas Eve |
| Dashing Away With The Smoothing Iron |
| Dashing White Sergeant |
| Darcy Farrow |
| Echo Canyon |
| Fanny Blair |
| Finuola, the Gem of the Roe |
| Four Walls |
| Heather on the Moor |
| Henry Martin |
| Henry, My Son |
| John Connolly, The Irish Rebel |
| King Henry, my Son |
| Kissin’s No Sin |
| Peggy Gordon |
| Queen Jane |
| The Arsenic Tragedy |
| The Ballad of John MacLean |
| Your Daughters and Your Sons |
| You Gentlemen of High Reknown |

# Curriculum Vitae

**Name:**

Ethan Axler

**Place and Year of Birth:**

Vancouver, BC, Canada, 2002

**Secondary School Diploma:**

King David High School, Vancouver Canada

**Experience:**

| | |
|---|---|
| Huron Auditory Perception Lab – Undergrad Research Assistant | 2020 – 2024 |
| BC Mental Health & Addiction Research Institute – Research Assistant | Summer 2022 |

**Awards:**

| | |
|---|---|
| CURL Poster Award | 2023 |
| Dean's Honours List | 2020/21 |
| Scholarship of Excellence | 2020 |
| Huron National Scholarship | 2020 |
| Valedictorian of Graduating Class of 2020, King David High School | 2020 |