12-1-2022

# The Stability of the Speech-to-Song Illusion and Individual Differences

Rodica R. Constantine

## Repository Citation

Constantine, Rodica R., "The Stability of the Speech-to-Song Illusion and Individual Differences" (2022). *UNLV Theses, Dissertations, Professional Papers, and Capstones*. 4581.
https://digitalscholarship.unlv.edu/thesesdissertations/4581

# THE STABILITY OF THE SPEECH-TO-SONG ILLUSION AND INDIVIDUAL DIFFERENCES

By

Rodica R. Constantine

Bachelor of Science – Psychology
Bachelor of Arts – Music
University of Pittsburgh
2018

A thesis submitted in partial fulfillment
of the requirements for the

Master of Arts – Psychological and Brain Sciences

Department of Psychology
College of Liberal Arts
The Graduate College

University of Nevada, Las Vegas
December 2022

UNLV | GRADUATE COLLEGE

# Thesis Approval

The Graduate College
The University of Nevada, Las Vegas

December 1, 2022

This thesis prepared by

Rodica R. Constantine

entitled

The Stability of the Speech-to-Song Illusion and Individual Differences

is approved in partial fulfillment of the requirements for the degree of

Master of Arts – Psychological and Brain Sciences
Department of Psychology

Erin Hannon, Ph.D.
*Examination Committee Chair*

Joel Snyder, Ph.D.
*Examination Committee Member*

Colleen Parks, Ph.D.
*Examination Committee Member*

Diego Vega, DMA
*Graduate College Faculty Representative*

Alyssa Crittenden, Ph.D.
*Vice Provost for Graduate Education &*
*Dean of the Graduate College*

## Abstract

Music and language are easily distinguishable for the average listener despite sharing many structural acoustic similarities. The Speech-to-Song illusion can give rise to both musical and linguistic percepts by inducing a perceptual switch after listening to multiple repetitions of a natural spoken utterance. As such, it has been used as a tool to control for low-level acoustic characteristics previously shown to drive lateralized brain responses regardless of domain-type, helping to disambiguate the contribution of high- versus low-level processes in both music and speech perception. However, there exists a lack of research on how large a role individual differences such as musical ability, tonal enculturation, sensitivity to speech prosody, and attention to lyrical content play in the elicitation and long-term stability of the Speech-to-Song illusion, which limits our understanding of how top-down musical and linguistic knowledge modulate perception. In our study, we measured the STS illusion by presenting listeners with excerpts known to elicit the STS illusion and asking them to rate the degree to which each repetition sounded song-like across delays from 0-56 days. To measure individual differences, we administered the Goldsmiths Musical Sophistication Index (Gold-MSI), a speech prosody test (PEPS-C), and a tonality test (from Corrigall & Trainor, 2015). Our results indicate the STS illusion increases in strength, is more readily elicited over delays, and also empirically validate anecdotal evidence that the STS illusion is temporally stable. Moreover, STS elicitation and consistency of STS excerpt ratings across sessions was predicted by many of our individual difference measures. This work holds important implications for understanding music and language processing, as well as memory for auditory stimuli.

*Keywords: Speech-to-Song Illusion, music, language, speech prosody, domain-specific*

# Acknowledgements

First and foremost, I thank my advisor, Dr. Erin Hannon, for her patience and guidance in helping me see this project completed. It is with the energy taken from our illuminating discussions and her enthusiasm for research that I am propelled toward future scientific endeavors.

I also thank my thesis committee – Dr. Colleen Parks, Dr. Joel Snyder, and Dr. Diego Vega – for their invaluable insights and suggestions, which substantially improved the content of this work.

My thanks extend to all members of the UNLV Auditory Cognition and Development and Auditory Cognitive Neuroscience laboratories who have in any way contributed to the development of this project, from its initial to final stages.

Thank you to my wonderful friends who have helped me along the progression of this project. To my dear friend, Maggie McMullin, I give my heartfelt thanks for enlivening my day-to-day life with humor, thoughtful discussions, and above all, for her unwavering support of me in anything I wish to pursue.

Last but not least, I express my gratitude to my family for their continued support throughout my life, particularly with regards to my scientific pursuits, and for igniting in me a reverence and devotion for music. Vă mulțumesc din suflet!

# Table of Contents

# List of Figures

# Chapter 1: Introduction

Speech and song are both complex, dynamic forms of auditory input that share many structural similarities but are nevertheless easily distinguishable by the average listener. Remarkably, when repeated multiple times, a short speech excerpt can give rise to a robust initial percept of speech that subsequently transforms into a percept of song—a phenomenon known as the Speech-to-Song (STS) illusion (Deutsch, Henthorn, & Lapidis, 2011). Although it is not entirely clear why and how listeners experience the STS illusion, prior research suggests it hinges on low-level acoustic characteristics, as well as individual difference factors such as musical aptitude and attention to speech prosody. Since STS stimuli can elicit both musical and linguistic percepts, they provide an opportunity to disentangle the role of high- versus low-level processes in both music and speech perception while taking individual differences into account. To our knowledge, there has not been any research conducted to assess the long-term stability of STS illusion, which limits our understanding of how stable music- and language- specific perceptual memories of STS stimuli remain over time. Anecdotally, after a listener has experienced the STS illusion with a given stimulus, they immediately perceive it as song upon reencountering it even without any repetition—even though repetition is typically required to experience the STS illusion. These reports suggest that the STS illusion is temporally stable, but this has yet to be empirically validated. In this study, we measured the stability of the STS illusion over time and the relation between individual differences and rating consistency of STS excerpts.

According to modular accounts of processing, there exist distinct neural mechanisms which encapsulate specialized knowledge for separate domains such as music and language (Chomsky, 1959; Cosmides & Tooby, 1994; Fodor, 1983; Karmiloff-Smith, 2009). Indeed, musical and linguistic stimuli have been shown to activate different hemispheres of the brain; when people listen to music, they show a right lateralized response, whereas when they listen to speech, they show a left lateralized response (Chen et al., 2021; Tervaniemi, & Hugdahl, 2003; Zatorre, Evans, Meyer, & Gjedde, 1992). In the past, lateralized brain responses to music and language have been interpreted as evidence of domain-specific language and music modules (Hébert, Racette, Gagnon, & Peretz, 2003; Peretz, 2001; Peretz & Hyde, 2003). More recent findings have demonstrated music- and language-specific activation in early processing areas (i.e., secondary auditory cortex), and a lack of activation to music stimuli in more widespread language networks (Chen et al., 2021; Fedorenko, McDermott, Norman-Haignere, & Kanwisher, 2012; Norman-Haignere, Kanwisher, & McDermott, 2015). There also exists evidence that low-level acoustic cues can drive lateralized responses regardless of stimulus domain due to the asymmetric sampling time of acoustic characteristics (Zatorre, Belin, & Penhune, 2002; Zatorre & Gandour, 2008). Stimuli with rapid temporal changes on the scale of 10s of ms have been shown to elicit responses in the left auditory cortex, whereas spectrally complex stimuli with slower temporal envelope changes on the scale of 100s of ms elicit a response in the right auditory cortex (Luo & Poeppel, 2012; Norman-Haignere et al., 2015; Poeppel, 2003; Rogalsky, Rong, Saberi, & Hickok, 2011; Zatorre et al., 2002). Given that speech is characterized by rapid temporal changes such as voice onset time differences and formant transitions, and music is predominated by sustained and discrete pitches, this evidence may suggest that at least some hemispheric lateralization is due to the acoustic properties of speech versus music (Joanisse & Gati, 2003; Zatorre et al., 2002).

These findings underscore the importance of controlling for low-level acoustic properties when trying to understand higher-level processing of music and language, since both can give rise to differential patterns of neural activity regardless of domain. As such, domain-specificity can be assumed when differential patterns of activity are elicited from a change in the high-level representation of a stimulus while its acoustic properties remain constant. This approach has been applied with sine-wave speech (SWS), a synthetic derivation of a natural utterance composed of multiple time-varying sinusoids that approximate the frequency and amplitude patterns of its resonance peaks (Remez, Rubin, Pisoni, & Carrell, 1981). SWS is first heard as meaningless sound and can later be perceived as a speech stimulus after some exposure or attentional cueing, which shows that a stimulus' low-level acoustic characteristics can be divorced from a listener's high-level perception of the same stimulus (Dehaene-Lambertz, Pallier, Serniclaes, Sprenger-Charolles, Jobert, Dehaene, 2005; Mottonen, Calvert, Jaaskelainen, Matthews, Thesen, Tuomainen, & Sams, 2006; Remez et al., 1981). The transition from hearing SWS as non-speech to hearing it as speech is accompanied by increased left-hemispheric activation (Dehaene-Lambertz et al., 2005; Remez et al., 1981). However, SWS has also been shown to elicit left-hemispheric activation compared to silence when heard as non-speech (Dehaene-Lambertz et al., 2005). This result is potentially driven by the speech-like properties preserved within non-speech such as formant frequency characteristics which may cause an increased attempt of the listener to identify sounds preferentially processed by the left-hemisphere. The STS illusion may provide a musical analogue to the SWS example because the same stimulus can elicit both speech and song percepts and thus disentangle low-level from high-level music and speech processing.

Previous studies have reported that the STS illusion is elicited by a range of speech stimuli (Deutsch, 2003; Deutsch et al., 2011; Leung & Zhou, 2018; Margulis, Simchy-Gross, & Black, 2015; Rowland, Kasdan, & Poeppel, 2019; Tierney, Dick, Deutsch, & Sereno, 2013;

Vanden Bosch der Nederlanden, Hannon, & Snyder, 2015a; Vanden Bosch der Nederlanden, Hannon & Snyder, 2015b), and even extends to non-speech stimuli (e.g., randomized tone sequences and environmental sounds; Deutsch et al., 2011; Margulis & Simchy-Gross, 2016; Simchy-Gross & Margulis, 2018; Tierney, Patel, & Breen, 2018). However, repetition alone does not always lead excerpts to transform to song; some speech excerpts seem to transform readily while others do not (Graber, Simchy-Gross, & Margulis, 2017; Tierney et al., 2013). Acoustic features that increase the likelihood a listener will experience the STS illusion include syllables with greater fundamental frequency (F0) stability and isochronous rhythmic patterns of stressed syllables (Falk, Rathcke, & Dalla Bella, 2014; Groenveld, Burgoyne, & Sadakata, 2020; Tierney et al., 2013; Tierney, Patel, & Breen, 2016).

Stimulus characteristics are not the only factor that predicts whether the STS illusion will be experienced; individual listener factors may also play a role. For example, in some studies the same stimuli that transform to song for some listeners remain speech-like for others (Vanden Bosch der Nederlanden et al., 2015b), though it is unclear why this is the case. Also, the frequency, strength, and amount of time needed to experience the STS illusion does not change as a function of age, between younger to older adults (Mullin, Norkey, Kodwani, Vitevitch, & Castro, 2021). Although musicians tend to rate stimuli as sounding more song-like than non-musicians, musicians are no more likely than non-musicians to experience the illusion (Mullin, et al., 2021; Deutsch et al., 2011; Falk et al., 2014; Vanden Bosch der Nederlanden et al., 2015b). Even though musicians outperform non-musicians at detecting pitch changes to speech excerpts (Falk & Rathcke, 2010; Vanden Bosch der Nederlanden, et al., 2015a), musicians may require more repetition than non-musicians to experience the STS illusion (Falk et al., 2014). This difference is possibly driven by musicians’ advanced encoding of musical versus prosodic speech representations which causes them to initially focus more on the speech-relevant acoustic properties of a stimulus longer than nonmusicians, in effect delaying the rate of STS illusion transformations (Besson, Chobert & Marie, 2011; Dalla Bella, Peretz, & Aronoff, 2003; Falk et al., 2014). Additionally, musicians' rich musical background could make for more refined selectivity of an auditory signal as speech- or song-like (Falk et al., 2014). Language background may also influence the likelihood of experiencing the illusion. Listeners are more likely to experience the STS illusion when presented with stimuli that are less "pronounceable" to them (e.g., excerpts spoken in a foreign language) (Margulis et al., 2015). The STS illusion has also been found to be weaker among speakers of tonal languages such as Mandarin Chinese, perhaps because they have a more well-defined template for how pitch functions within a linguistic context and are therefore better able to resist interpreting pitch within language as musical (Bidelman & Lee, 2015; Jaisin, Suphanchaimat, Figueroa Candia, & Warren, 2016). Thus, while musical and linguistic experience may influence whether or not an individual experiences the STS illusion, these individual differences remain largely unexplained.

Research investigating the factors that give rise to the STS illusion is important for explaining the underlying cognitive mechanisms enabling the perceptual switch from hearing the same stimulus as speech to hearing it as song. One proposal is that highly repetitive forms of language shift one's attention away from the syntactic and semantic features of a stimulus to aspects that are characteristically musical, such as individual pitches and temporal regularities (Graber et al., 2017; Margulis, 2014). In the case of the STS illusion, the exact same, unmodified repetition of a speech excerpt is needed to facilitate the perceptual switch from a speech to a music listening mode (Deutsch et al., 2011; Graber et al., 2017; Margulis, 2014). Node Structure Theory (NST) emphasizes the role of semantic satiation, a phenomenon in which the repetition of a word or phrase can cause it to temporarily lose its meaning (Castro, Mendoza, Tampke, & Vitevitch, 2018; MacKay, 1987; Severance & Washburn, 1907). According to this connectionist model, repetition leads to the satiation of lexical nodes which results in the temporary loss of higher-level semantic content of the excerpt, allowing for the continual priming and increased salience of speech rhythmic properties (Castro et al., 2018; Cutler, 1991; Jackendoff, 2009). As the stimulus repeats, the listener begins to perceive syllables as more beat-like, which induces the STS illusion (Castro, 2014; Castro et al., 2018; Vitevitch, Ng, Hatley, & Castro, 2021). Rate of semantic satiation allowing for STS transformation might also be contingent on the clustering coefficient of phonological neighborhoods (i.e., the extent to which neighbors of a word are also neighbors of each other; Vitevitch, et al., 2021). Listeners tend to reactivate a priming mechanism at the lexical node faster when the clustering coefficient of phonological neighborhoods is high, giving rise to semantic processing, which entails lower song-like ratings. Conversely, when the clustering coefficient of phonological neighborhoods is low, priming is distributed to syllable nodes, giving rise to a stronger song-like percept (Vitevitch et al., 2021). Falk and colleagues (2014) offer a similar interpretation, namely that through repetition, the listener engages in the continual reinterpretation of a given speech excerpt's prosodic characteristics, motivating the switch from linguistic to musical processing (Falk et al., 2014; Tierney et al., 2013). The significant effect of exact repetition has also been shown at the level of the speaker, where speech excerpts repeated by the same – as opposed to different – speakers are perceived to be more song-like, potentially because semantic satiation can occur more readily when acoustic characteristics are held constant (Soehlke, Kamat, Castro, & Vitevitch, 2022). Although these accounts all emphasize the role of repetition in reducing the salience of speech meaning, they are less clear about what a “musical mode” of listening entails or how individuals might apply their musical experience and knowledge when experiencing the illusion.

Music-specific representations are implicated in the STS illusion in several ways. In the original STS illusion experiment, Deutsch et al. (2011) asked musicians to vocally reproduce a speech excerpt after repetitions one and ten. While participants’ first reproduction sounded like speech, by the tenth repetition, they subtly adjusted the excerpt’s overall pitch contour to be more in line with a Western musical scale. Although Deutsch et al. (2011) interpreted these ostensibly musical alterations to be indicative of music-specific processing, they could have been an effect of using a vocal reproduction task, since musicians are trained to sing in tune with the musical scale. However, using a purely perceptual task, Vanden Bosch der Nederlanden et al. (2015b) provided evidence that participants recruit music-specific knowledge when making judgements about pitch modifications to stimuli that transform to song. Participants in their study had an advantage for detecting “non-conforming” pitch changes that violated the Western musical scale over “conforming” pitch changes that were consistent with listeners’ presumed musical expectations (see Figure 1). They only showed this advantage, however, after stimuli had transformed to song, and not for stimuli that were always perceived as speech. This suggests that for adult listeners, the same stimulus can give rise to speech percepts that give rise to linguistic expectations as well as musical percepts that entail distinct expectations about scale, key, and harmony. Above all, both experiments showed that when participants experience the STS illusion, they assimilate pitch contours in speech to a Western musical scale, providing evidence that they are engaging implicit musical knowledge when they perceive speech excerpts as song.

Although musicians and non-musicians both experience the STS illusion (Vanden Bosch der Nederlanden et al., 2015a), it remains unclear how much individual differences in music-specific knowledge may influence a listener's perception when an excerpt transforms to song. One distinguishing advantage of having musical experience seems to be an increased sensitivity to pitch which allows for making fine musical discriminations such as noticing changes in pitch interval sizes (Besson, Schön, Moreno, Santos, & Magne, 2007; Schön, Magne, & Besson, 2004). Musicians have also been shown to outperform non-musicians in working and short-term memory pitch discrimination tasks (Dowling, 1991; Dowling, Bartlett, Halpern, & Andrews, 2008), demonstrating an enhanced ability to sustain cognitive control—a possible effect of previous focused musical training (Pallesen, Brattico, Bailey, Korvenoja, Koivisto, Gjedde, & Carlson, 2010; Zuk, Benjamin, Kenyon, & Gaab, 2014). Both musicians and non-musicians exhibit robust tonal expectations (Fujioka, Trainor, Ross, Kakigi, & Pantev, 2005; Marmel, Tillmann, & Dowling, 2008). However, while non-musicians rely on implicit musical knowledge, musicians rely on both implicit and explicit musical knowledge they acquired through training and have shown superior auditory recognition memory for musical and speech stimuli relative to nonmusicians (Cohen, Evans, Horowitz, & Wolfe, 2011; Cohen, Horowitz, & Wolfe, 2009; Tillmann, 2005). Hence, musicians' tonal knowledge might give rise to a stronger assimilation of a speech excerpt to the musical scale once it transforms to song (Dowling, 1991; Guo & Koelsch, 2016; Hansen, Vuust, & Pearce, 2016; Marmel & Tillmann, 2009; Park, Chung, Kim, Lee, Seol, & Yi, 2018). We might therefore expect that musicians would form stronger musical representations of transformed speech excerpts, and thus better detect pitch changes that violate tonal expectations. Conversely, musicians could be less affected by tonal expectations during a discrimination task because they have been trained to detect pitch changes regardless of key or scale, whereas non-musicians' expectations rely more on implicit knowledge gained from everyday listening experiences which do not entail making such distinctions. A final possibility is that musicians and non-musicians perform similarly, assuming the STS illusion primarily activates implicit musical knowledge in both groups.

It is currently unclear how individual differences might affect the long-term stability of the STS illusion. Past research has shown that, relative to atonal melodies, tonal melodies conforming to Western listeners' musical schemas are preferentially consolidated during REM – as opposed to slow wave—sleep over a 24-hour window (Durrant, Cairney, McDermott, & Lewis, 2015), suggesting that when activated, musical schemas for tonal melodies enhance memory. Musical properties of STS stimuli which activate such tonal representations could be predictive of their long-term stability, as representations of absolute musical properties like pitch and tempo have shown some stability across delays (Frieler, Fischinger, Schlemmer, Lothwesen, Jakubowski, & Müllensiefen, 2013; Levitin, 1994; Schellenberg & Habashi, 2015). Having stronger musical representations of STS stimuli might therefore predict long-term stability of the STS illusion regardless of training, where those who show the strongest effects of conforming over non-conforming pitch changes might also exhibit the most stable STS illusions, while those who show the weakest effects could exhibit weaker STS illusion stability. Using a "gating" paradigm in which novel and familiar melody lines are presented one note at a time to gauge the rate of melody recognition, Dalla Bella et al. (2003) found that musicians exhibit faster recognition of familiar melodies as compared to non-musicians. This finding could suggest that if the STS illusion is stable across delays, those with more musical training would recognize stimuli which previously transformed to song as song-like more readily upon reencountering them. However, the rate of memory decay for both musicians and non-musicians is remarkably similar when making recency and frequency judgements for novel and familiar stimuli in long-term memory (Kauffman & Carlsen, 1989; McAuley, Stevens, & Humphreys, 2004), which might predict that the long-term stability of the STS illusion would be similar for musicians and non-musicians.

Experiencing the STS illusion is contingent on the propensity of the listener to interpret a speech stimulus as musical, which suggests that tuning in to the musical properties of speech (i.e., speech prosody) is critical to experience the illusion (Boutsen, 2003). Not only does native language tonality predict how listeners attend to pitch information in speech (Bidelman & Lee, 2015; Jaisin et al., 2016), sensitivity to speech prosody might also correlate with musical aptitude. For instance, listeners with more musical training are better able to judge emotion in naturalistic excerpts of music and speech (Dibben, Coutinho, Vilar, & Estévez-Pérez, 2018). Musical training also appears to enhance early and later stages of semantic and pure prosodic processing (i.e., unintelligible semantic content) (Pinheiro, Vasconcelos, Dias, Arrais, & Gonçalves, 2015). Moreover, musical aptitude has been linked to elevated processing of pitch violations in music and language by musician children as compared to non-musician children, showing potential advantages of musical training in the ability to process pitch information from both domains (Magne, Schön, & Besson, 2006).

On the other hand, those with deficient music processing (such as adults with congenital amusia) exhibit reduced sensitivity to emotional speech prosody and elevated thresholds for pitch-change detection in both musical and linguistic stimuli (Liu, Patel, Fourcin, & Stewart, 2010; Lu, Ho, Liu, Wu, & Thompson, 2015; Patel, Foxton, & Griffiths, 2005; Thompson, Marin, & Stewart, 2012). Collectively, these findings suggest there could be overlapping pitch processing for music and language, where better pitch processing for music also predicts better pitch processing for language. However, the causal relationship between musical training and prosodic processing enhancements remains unclear. It could be that musical training improves prosodic processing, or that individuals who naturally tune in to musical properties of language (i.e., “musical listeners”) are more likely to pursue musical training. In the latter case, it is possible that there exists a subset of individuals who are adept at prosodic processing but did not happen to pursue musical training, underscoring the need to study both musical training and prosodic processing. Given that the perceptual switch from speech to song is thought to occur when the listener directs their focus away from the semantic content of an excerpt and onto its musical aspects conveyed through speech prosody, listeners who naturally attune to the prosodic aspects of speech most likely activate strong musical representations and thus experience the STS illusion more readily and robustly than those less adept at prosodic processing, which might also entail greater long-term stability of the STS illusion. Regardless, it remains unclear the extent to which prosodic processing is divorced from semantic content once a speech excerpt transforms to song.

This study examined how STS illusion elicitation and consistency of participant ratings of STS excerpts over time is influenced by individual differences in various measures of musicality and sensitivity to speech prosody. We first measured the long-term stability of the STS illusion over delays of up to 56 days to empirically test anecdotal reports that it is stable. Then, we examined how mentioned individual differences contribute to the experience and consistency of STS excerpt ratings across sessions.

# Chapter 2: Method

The primary goal of this experiment was to assess the stability of the “musical mode” of processing that may persist after listening to particular utterances that exhibit the STS illusion by testing participants across two sessions separated by hours, one day, 7 days, 28 days, or 56 days. If anecdotal reports hold, participants who heard a given stimulus transform to song during their first session will continue to hear that stimulus as song at the second session, after fewer or no repetitions. Moreover, assuming the STS illusion engages music-specific representations of musical scale and key, we expect listeners to better detect non-conforming than conforming pitch changes for stimuli heard as song. Our secondary goal was to examine individual variation in STS illusion elicitation and consistency of STS excerpt ratings. Specifically, we asked whether individual differences in musical sophistication, tonal enculturation, sensitivity to prosodic speech cues, and attention to lyrical content predict the consistency of ratings given to STS excerpts across sessions, as well as the accuracy with which listeners detect pitch changes that violate tonal expectations.

## Participants

One hundred and fifty-eight participants were recruited from the undergraduate psychology subject pool at the University of Nevada, Las Vegas (UNLV) along with volunteers from across the United States. Participants from the UNLV subject pool were granted course credit. Participants were required to be English-speaking, over 18 years of age, and have no known hearing deficits or uncorrected visual deficits.

### Stimuli

For both the rating and discrimination tasks, we used 24 natural speech excerpts from open-access audiobook recording websites of three male speakers, from Tierney et al. (2013) (see all of our project materials here: https://osf.io/qdpcf/). These excerpts were previously reported to give rise to a robust STS illusion, presumably because of their relatively greater fundamental frequency stability and greater regularity of inter-stress intervals between syllables compared with excerpts that did not transform to song (Tierney et al., 2012). For the rating task practice trial, we used the "Sometimes behave so strangely" excerpt provided by Deutsch et al. (2011).

For the discrimination task, change stimuli were created by making subtle pitch modifications to the above stimuli. We used two types of changes: "conforming," which conform to presumed Western tonal expectations, and "non-conforming," which violate Western tonal expectations. To create these pitch modifications, we used musical transcriptions of speech stimuli, singing, and experimenter intuition to approximate each speech excerpt's sung contour after it undergoes the STS illusion. Next, target syllables were selected based on the following criteria: 1) a stable fundamental frequency without pitch glides, as pitch glides could make the discrimination task too difficult and 2) the spoken syllable is either sharp or flat compared to its sung contour by 0.2-1.0 semitone. Syllables that met these criteria were then raised or lowered by a 20-90% semitone shift to make the speech excerpt match or deviate from musical expectations. Conforming pitch changes were moved toward the sung contour whereas non-conforming pitch changes moved away from the sung contour by the same amount. For example, if the fundamental frequency of a syllable is 192.97 Hz (G3 minus 27 cents on the musical scale) and sounds flatter than its sung version by 0.3 semitones, the conforming change would raise the syllable pitch to match musical expectation by 0.3 semitones to 196.86 Hz (G3 plus 8 cents), whereas the non-conforming change would lower it by 0.3 semitones to 190.28 Hz (F#3 plus 49 cents) (see Figure 1).

All modifications were made in Praat and excerpts were exported using the PSOLA overlap-add method to preserve formant frequency characteristics (Praat Software; Boersma & Weenink, 2001; Vanden Bosch der Nederlanden, et al., 2015b). For the discrimination task practice trial, we manipulated the “Sometimes behave so strangely” stimulus provided by Deutsch et al. (2011) to include pitch changes for both the conforming and non-conforming conditions.

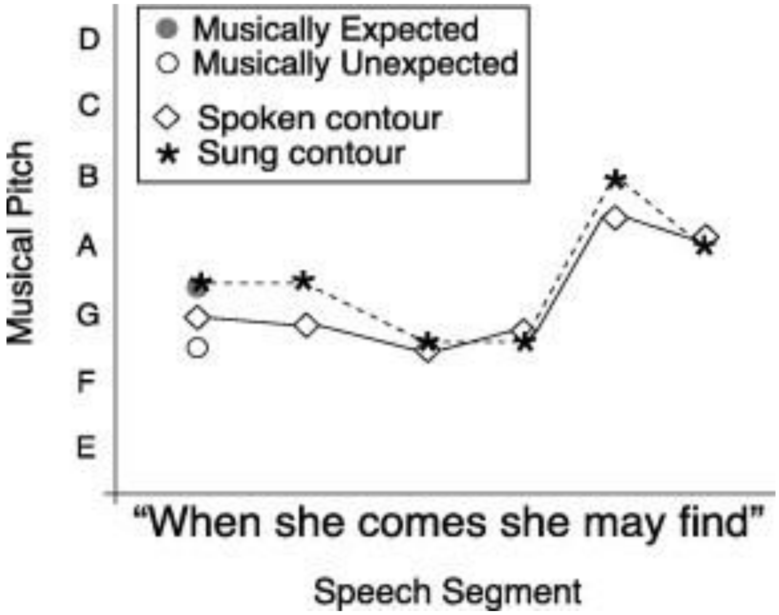

**Figure 1.** *This figure depicts both the musically conforming and non-conforming changes made to the syllable "When" within the speech excerpt "When she comes she may find." Here, open diamonds represent the approximate fundamental pitch stability of each spoken syllable, while the asterisks exemplify what would be the sung contour of each pitch through the lens of the Western musical scale. The closed, gray circle represents a pitch change of the spoken syllable within the musically expected condition conforming to the sung contour, and the open circle represents a pitch change of the same syllable in the musically unexpected condition which deviates from the sung contour. Both expected and unexpected pitch changes were matched in distance from the approximate pitch the spoken syllable. Note that rhythmic information is not illustrated in this figure (Vanden Bosch der Nederlanden, et al., 2015b).*

### Procedure

This two-session experiment was conducted through the online survey tool, Qualtrics (Qualtrics, Provo, UT), with custom HTML script to play audio. Informed consent was obtained on-line from each participant prior to beginning the study. Participants were randomly assigned to one of five delay conditions: same day (N = 32), one day (N = 31), 7 days (N = 35), 28 days (N = 30), or 56 days  (N = 30). For the same day delay condition, participants took a short break of approximately 5 minutes between session 1 and session 2. For all other delay conditions, participants received e-mail reminders to reserve a time slot within 48-hour time window during which they are expected to participate in session 2. Since a subset participants took the study online instead of in the lab, a Qualtrics filter was added for additional control which restricted participants to using a computer instead of a smart phone or tablet. Participants were required to wear headphones and asked to adjust their volume to a comfortable listening level. Session 1 consisted of 1) a headphone check, 2) the rating task, 3) the discrimination task, 4) individual difference measures, 5) a compliance and attention check. Session 2 consisted of 1) a headphone check, 2) the rating task, 3) the discrimination task, 4) a compliance and attention check.

### *Headphone Check*

To ensure quality of on-line data collection, we tested the sound quality heard by each participant by administering a headphone check created and validated by Woods, Siegel, Traer, & McDermott (2017). The headphone check includes 6 trials of a 3-AFC intensity discrimination task for which participants must indicate which tone was the softest by selecting one of three buttons labeled "Tone 1", "Tone 2", or "Tone 3." Any participant who did not correctly answer 5 of 6 trials were excluded (Woods et. al, 2017).

### *Rating Task*

For the rating task, participants heard each of the 24 stimulus excerpts repeat 10 times per trial, and after each repetition they were asked to indicate "How song-like or speech-like does this excerpt sound?" on a scale of 1 (exactly like speech) to 5 (exactly like singing) via mouse click. The time in which participants responded was not limited, but they were required to enter a rating of each repetition to advance to the next. The arrangement of the 24 stimuli varied across unique 9 lists. Prior to the test trials, participants were introduced to one practice trial in which they heard the speech excerpt "sometimes behave so strangely" repeat 10 times and provided a rating for each repetition (Deutsch et al., 2011). For both sessions, each excerpt was repeated 10 times per trial to replicate previous research (Deutsch et al., 2011); however, prior research suggests that, on average, most speech excerpts transform around the third repetition (Falk et al., 2014).

### *Discrimination Task*

At each session, participants performed the discrimination task with each of the 24 speech excerpts used in the rating task. On each trial, a standard stimulus was presented twice followed by a comparison stimulus, and participants were asked to indicate whether the comparison stimulus was the same or different from the standard stimulus by selecting the option “same” or “different” via mouse click. On “no change” trials, the comparison stimulus was identical to the standard stimulus. On change trials, the comparison stimulus contained pitch changes that either conformed or did not conform to Western musical structure (see Figure 1). Each of the 24 excerpts were presented across 24 trials, divided into 8 same trials, 8 non-conforming change trials, and 8 conforming change trials. Nine discrimination task trial orders were designed to ensure all three stimulus conditions were counter-balanced for each speech excerpt in a between-subjects fashion. Before beginning test trials, participants were given a practice trial of the speech excerpt “sometimes behave so strangely” which contained a conforming or non-conforming pitch change.

### *Individual Difference Measures*

At session 1, participants were asked to complete the Gold-MSI version 1.0 (Müllensiefen, Gingras, Musil, & Stewart, 2014), a questionnaire featuring five subscales (Active Musical Engagement, Perceptual Ability, Musical Training, Singing Ability, and General

Musical Sophistication) to measure individual differences in perceived musicality. We also measured how closely participants pay attention to lyrical content, by their response to the question “Some people listen to music and pay close attention to the lyrics and the message, whereas other people focus on the feeling and sounds of the music to the point they are sometimes surprised when they learn what a familiar song is about. How closely would you say you attend to the lyrics while listening to music?” on a Likert scale ranging from 1 to 7 (1 = not closely, 7 = very closely).

At session 1, participants also completed a set of music and language perceptual tests. For the music perception test, we reproduced a melodic discrimination task by Trainor and Trehub (1994) to test participants’ tonal expectations. For this task, we composed four melody lines according to the criteria set forth in Trainor and Trehub (1994), where a pitch change was made to the sixth note of each 10-note melody to create three conditions per melody: out-of-key, out-of-harmony, and in-harmony. The five melodies were counterbalanced across nine lists in three different keys (C major, E major, and A-flat major), where participants received two melodies per condition, including melodies for which no change was made, totaling eight trials. On each trial, a standard stimulus was presented twice followed by a comparison stimulus, and participants were asked to indicate whether the comparison stimulus was the same as or different than the standard stimulus by selecting the option “same” or “different” via mouse click. On “no change” trials, the comparison stimulus was identical to the standard stimulus.

Then, we administered the Profiling Elements of Prosody in Speech Communication (PEPS-C; Peppé & McCann, 2003), which measured participants’ ability to interpret prosodic cues within speech across 3 iterations of each of 6 subtasks (Turn End, Affect, Lexical Stress, Phrase Stress, Boundary, and Contrastive Stress).

### *Compliance and Attention Checks*

At the end of each session used four questions, modified for our study from Mehr, Singh, York, Glowacki, & Krasnow, 2018 to assess compliance and attention: 1) “What color is the sky? Please answer this incorrectly, on purpose, by choosing YELLOW instead of blue”, 2) “People are working on this task in many different places. Please tell us about the place you worked on this task. Please answer honestly.”, 3) “Did you wear headphones while listening to the sounds in this experiment? Please answer honestly. ”, 4) “Students are working on this task with many different devices, browsers, and internet connections. Please tell us about whether you had difficulty loading the sounds. Please answer honestly.”, and 5) “How carefully did you complete this survey? Please answer honestly.” (For response options and exclusion criteria, see Appendix A). 352 participants were excluded for failing the headphone and/or compliance and attention checks at either session.

# Chapter 3: Results

## Rating Task Analyses

To identify which stimuli were perceived as speech or song during session 1 (“Initial Percept”), for each session 1 trial we calculated trial-final averages of the ratings (“How song-like or speech-like does this excerpt sound?” on a scale of 1 = most like speech to 5 = most like song) for the last two repetitions (i.e., repetitions 9 and 10). Excerpts with trial-final ratings below 3 at session 1 were categorized as “speech” and ratings of 3 or higher were categorized as “song.”

To examine the stability of the STS illusion, we first asked whether excerpts classified at session 1 as “song” were still rated higher (more song-like) at session 2 than those initially classified as “speech.” We submitted trial-final session 2 ratings to a 2 x 5 mixed design ANOVA with Initial Percept [speech, song] as a within-subjects factor and delay condition (Delay [0, 1, 7, 28, or 56 days]) as the between-subjects factor. There was a statistically significant main effect of Initial Percept across delays $F(1, 150) = 652.16$, $p < .001$, $\eta_p^2 = 0.81$, where at session 2, participants gave higher ratings (i.e., more song-like) to stimuli they initially rated as song at session 1 (M = 3.72, SD = 0.61) than stimuli they initially rated as speech (*M* = 2.23, SD = .73). There was no significant main effect of Delay $F(4, 150) = 1.02$, $p < .399$, $\eta_p^2 = 0.03$, and no significant interaction between Initial Percept and Delay $F(4, 150) = 0.73$, $p = 0.57$, $\eta_p^2 = 0.02$. These results indicate that a participant’s initial categorization of an STS stimulus as speech or song persists across sessions, regardless of delay condition (Figure 2).

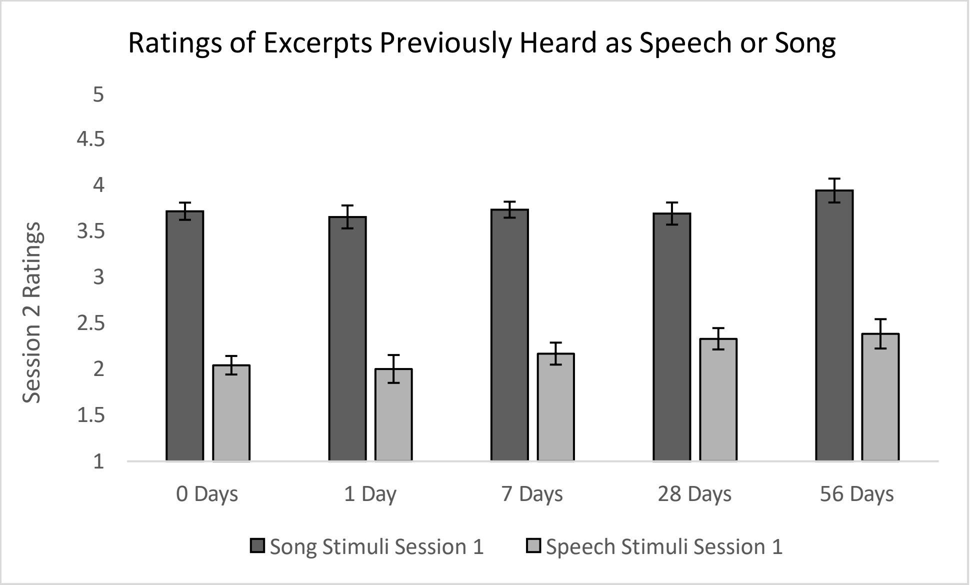

**Figure 2.** At session 2, stimuli initially rated as "song" during session 1 were rated significantly higher than those rated as initially "speech."

To assess consistency of STS excerpt ratings from session 1 to session 2, we conducted cross-session correlations for each participant to examine the extent to which participants consistently rated each of the 24 stimuli across sessions. Of 158 participants, 132 had significant ($p < .05$), positive correlations between session 1 and session 2 ratings (*M* cross-session correlation = 0.62). To measure the effect of delay on cross-session correlations, we submitted individual participant correlations to a one-way ANOVA with Delay [0, 1, 7, 28, or 56 days] as a between-subjects factor. Cross-session correlations did not differ by delay $F(4, 154) = 0.71$, $p = 0.59$, $\eta_p^2 = 0.02$ (Table 1). Taken together, these results suggest ratings of STS illusion excerpts are highly stable across delays of 0 – 56 days.

**Table 1.**
*Mean, Range, and Standard Deviation, of Cross-Session Correlations*

| Cross-Session Correlations | *M (Range)* | *SD* |
|---|---|---|
| Delay 1 (0 days) | 0.63 (.15-.91) | 0.18 |
| Delay 2 (1 day) | 0.61 (-.19-.85) | 0.24 |
| Delay 3 (7 days) | 0.66 (.11-.93) | 0.19 |
| Delay 4 (28 days) | 0.60 (.16-.87) | 0.23 |
| Delay 5 (56 days) | 0.58 (-.06-.90) | 0.21 |

Given that the STS illusion arises due to repetition and exposure, we also asked whether or not the robustness of the illusion increased over the two sessions. Anecdotal evidence suggests that once a stimulus perceptually transforms to song, that stimulus will later be heard as song with minimal or no repetition. If this is indeed the case, we would predict stimuli that perceptually transform to song at an initial session would require fewer repetitions to be perceived as song at a later session. We therefore examined the number of repetitions necessary for a stimulus to perceptually transform from speech to song over the course of each trial across both sessions. We calculated the variable "transformation position," as the point during the trial (e.g., the repetition number) when subjects give a stable rating of 3 or higher. Note that this analysis only includes stimuli that were heard to perceptually transform to song at both sessions (roughly 52% of trials in session 1 and 58% in session 2, since excerpts that did not perceptually transform to song could have no transformation position). We submitted each participant's mean Transformation Position to a 2 x 5 mixed design ANOVA with Session [1, 2] as a within-subjects factor and Delay [0, 1, 7, 28, or 56 days] as the between-subjects factor. There was a statistically significant main effect of session across delays $F(1, 153) = 169.37$, $p < .001$, $\eta_p^2 = 0.53$, where stimuli transformed to song earlier at session 2 ($M = 2.32$) than at session 1 ($M =$

3.38). There was no significant interaction between session and Delay $F(4, 153) = 0.42$, $p > .05$, $\eta_p^2 = 0.01$, and no significant effect of Delay $F(1, 153) = 0.75$, $p > .05$, $\eta_p^2 = 0.02$. This result provides further support for the notion that once a stimulus is perceived to transform to song, it requires minimal repetition to again be heard as song up to 56 days later (Figure 3).

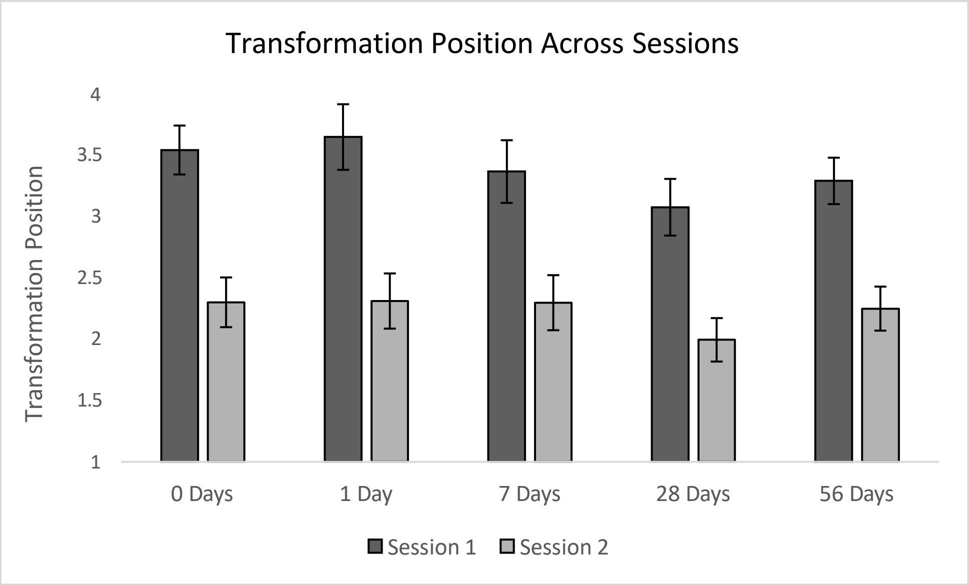

**Figure 3.** Transformation Position decreased from session 1 to session 2 across all delays.

Not only did stimuli transform to song earlier, but trial-final ratings for all stimuli were higher at session 2 compared to session 1. When we submitted trial-final ratings for all stimuli to a 2 x 5 mixed design ANOVA with Session [1, 2] as a within-subjects factor and Delay [0, 1, 7, 28, or 56 days] as the between-subjects factor, we observed a significant main effect of session, with ratings increasing slightly from session 1 ($M = 2.79$, SD = .73) to session 2 ($M = 2.97$, SD = .79) $F(1, 153) = 32.61$, $p < .001$, $\eta_p^2 = 0.18$.

There was no main effect or interaction with Delay (p > .05). We also examined whether more stimuli transformed to song at session 2 than at session 1, by calculating for each subject the total number (out of 24) of stimuli perceived to transform to song (e.g., the trial-final averages were 3 or greater). This variable was submitted to a 2 x 5 mixed design ANOVA with Session [1, 2] as a within-subjects factor and Delay [0, 1, 7, 28, or 56 days] as the between-subjects factor, which revealed a significant main effect of Session $F(1, 153) = 32.86$, $p < .001$, $\eta_p^2 = 0.18$, with more excerpts heard as song at session 2 ($M = 13.90$ ) than at session 1 ($M = 12.38$). There was no significant main effect or interaction with Delay (p > .05). This suggests that not only do participants continue to experience the illusion with excerpts they initially heard as song, but when repeating the task at a second session, excerpts initially heard as speech are more likely to transform to song. Together, these results suggest that the STS illusion is not only stable but may become more robust across repeated testing (Figure 4).

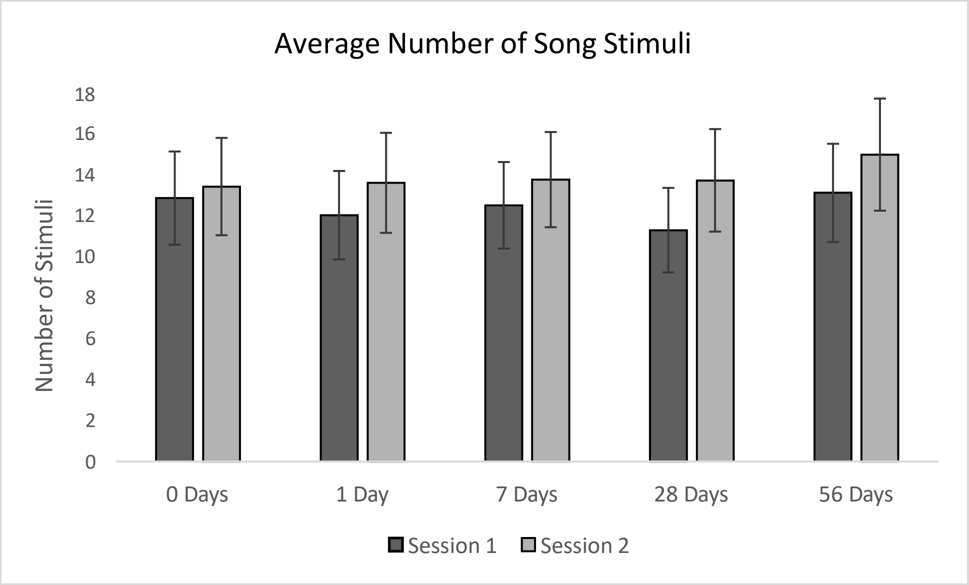

**Figure 4.** Mean number of excerpts (out of 24) heard as song at sessions 1 and 2.

## Discrimination Task Analyses

Next, we investigated the effects of the STS illusion on detection of pitch changes that conformed or violated Western musical scale structure. For each participant in each condition (musically conforming and musically non-conforming), discrimination (d') scores were calculated from proportion of hits (''different'' responses on change trials) and false alarms (''different'' responses on same trials) (d' = z(H) - z(F)). To account for extreme scores, we used the loglinear approach, for which the percent of signal trials to noise trials were respectively added to the numerator, and 1 was added to the denominator of total hits and false alarms per condition (Hautus, 1995; Stanislaw & Todorov, 1999). Based on prior work using the same excerpts and similar pitch changes (Vanden Bosch der Nederlanden, et al., 2015) we expected that pitch changes that violated Western scale structure would be easier to detect than changes that conformed to scale structure, particularly for excerpts that were heard to transform to song. Session 1 d' scores were submitted to a three-way 2 x 2 x 5 mixed design ANOVA with Initial Percept [speech, song] and Change Type [conforming, non-conforming] as within-subjects factors and Delay [0, 1, 7, 30, and 60 days] as the between-subjects factor. There was a main effect of Change Type $F(1, 157) = 11.73$, $p < .001$, $\eta_p^2 = 0.07$, such that d' was higher for incongruent ($M = 0.50$) than for congruent changes ($M = .34$). There was no significant main effect or interaction with Initial Percept (p > .05). We also performed this analysis with session 2 d' scores, and again observed a main effect of Change Type $F(1, 157) = 25.02$, $p < .001$, $\eta_p^2 = .14$ in favor of incongruent pitch changes ($M = 0.55$) over congruent changes ($M = 0.34$), and no significant main effect or interaction with Initial Percept (p > .05). These results replicate prior findings (Vanden Bosch der Nederlanden, et al., 2015b) that non-conforming changes were easier to detect than conforming changes, however unlike prior work we found this trend for all excerpts, not just excerpts that transformed to song (Figure 5).

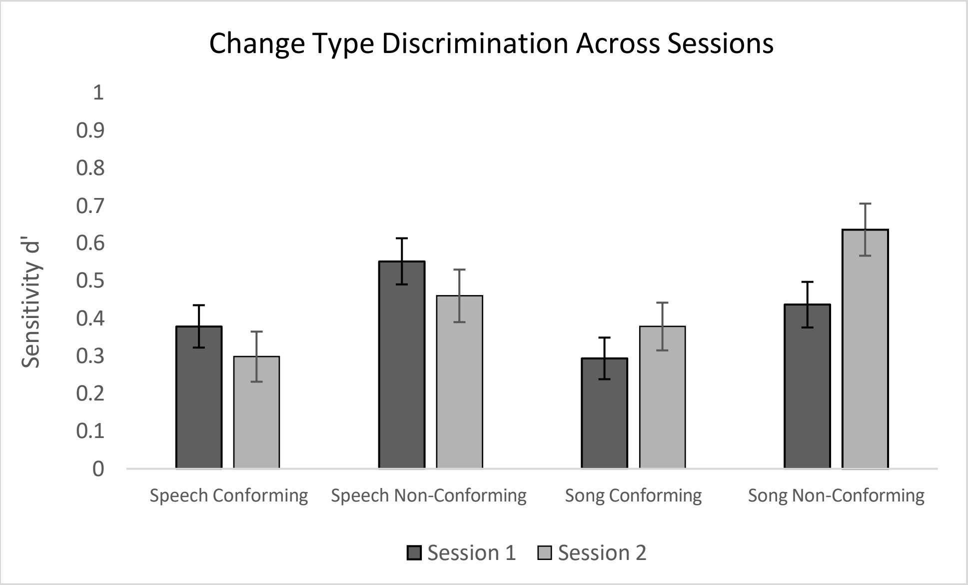

**Figure 5.** Non-conforming pitch changes were easier to detect than conforming pitch changes for both speech and song stimuli across sessions.

We also asked whether discrimination varied across sessions and according to delay. We calculated a cross-session change score by subtracting session 1 d’ scores from session 2 d’ scores (such that positive values reflect improvement across sessions) and submitted this variable to a 2 x 2 x 5 mixed design ANOVA with Initial Percept [speech, song] and Change Type [conforming, non-conforming] as within-subjects factors and Delay [0, 1, 7, 30, and 60 days]) as a between-subjects factor. There was a significant main effect of Initial Percept $F(1, 153) = 5.30$, $p < .05$, $\eta_p^2 = .03$, such that the increase in d’ was larger for excerpts initially perceived to transform to song ($M_{session1} = 0.19$ , $M_{session2} = -0.07$) than for those initially perceived as speech ($M_{session1} = 0.38$ , $M_{session2} = 0.31$) (see Figure 6). There was no significant main effect of Change

Type or interactions between Change Type and Initial Percept or Delay (p's > .05). However, there was a significant main effect of Delay $F(4, 153) = 2.66$, $p < .05$, $\eta_p^2 = .07$, $M = .12$, and post hoc comparisons with a Bonferroni correction indicated there was a significant difference between the 7 ($M = 0.49$) and 56 ($M = -0.01$) day delays (Mean Difference = 0.41 units, $p < .05$, 95% C.I. = [.03, .79]).

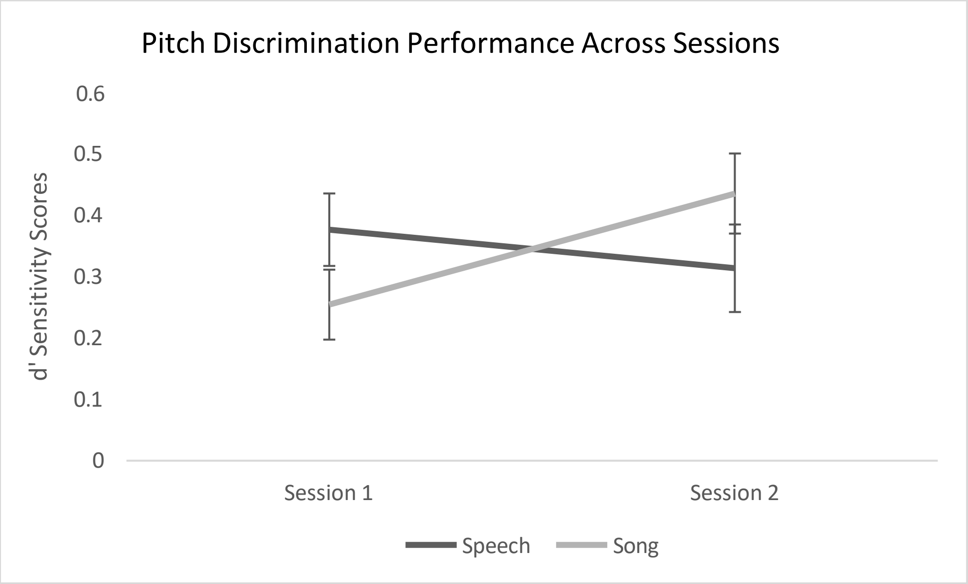

**Figure 6.** Mean d' detection for speech and song across sessions.

## Individual Difference Analyses

We then examined whether experience of the STS illusion was predicted by individual differences in musical sophistication (as measured using the Gold-MSI), attention to lyrical content, tonal enculturation (Trainor), and sensitivity to speech prosody. As described above (p. 17), the total Gold-MSI index score was obtained by averaging across five sub-scales (Active Musical Engagement, Perceptual Ability, Musical Training, Singing Ability, and Sophisticated Engagement) which combine self-report questions about musical aptitude, ability, engagement, and emotional responsiveness to music. To assess tonal enculturation, we calculated percent correct scores of participants' ability to detect out-of-key, out-of-harmony, and in-harmony pitch changes made to melodies in our tonal enculturation task (Trainor & Trehub, 1994). To attain a total tonal enculturation score, we calculated an average of percent correct of participant performance across conditions. Participants' sensitivity to prosodic speech cues was measured by calculating percent correct across all 6 subtasks (Turn End, Affect, Lexical Stress, Phrase Stress, Boundary, and Contrastive Stress) used in the "Profiling Elements of Prosody in Speech Communication" PEPS-C test of prosodic ability (Peppé & McCann, 2003). Finally, to gauge how closely participants pay attention to lyrical content, we used each participant's response to the question "How closely would you say you attend to the lyrics while listening to music?" on a Likert scale ranging from 1 to 7 (1 = not closely, 7 = very closely). As shown in Table 2, total Gold-MSI scores were significantly positively correlated with tonal enculturation and attention to lyrics, indicating participants who performed well on our objective tonality task also had a high self-report score of musical sophistication (r = 0.22) and reported paying more attention to lyrics (r = 0.24). Additionally, attention to lyrics was negatively correlated with tonal enculturation (r = -0.18) and prosodic sensitivity (r = -0.25), which may arise because participants who pay relatively less attention to lyrical content may instead focus on musical content including sounds, texture, and the general feeling music invokes for them.

**Table 2.**
*Summary of Means, Standard Deviation, and Correlations Among Individual Difference Factors*

| Variable | *M* | SD | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|---|
| 1. Total Gold-MSI Score | 3.99 | 1.16 | — | | | |
| 2. Tonal Enculturation | 57.04 | 15.98 | 0.22** | — | | |
| 3. PEPS-C (Prosody) | 88.60 | 7.30 | 0.12 | -0.01 | — | |
| 4. Attention to Lyrics | 4.63 | 1.61 | 0.24** | -0.18* | -0.25* | — |

* p < .05, ** p < .01, *** p < .001

We also examined correlations between predictor variables and several measures of the STS illusion. First, we asked how individual musical and prosodic abilities predict the initial experience of the STS illusion. We examined four dependent measures from session 1: 1) average ratings (based on trial-final ratings), 2) number of excerpts heard as song, 3) transformation position, and 4) overall d' for excerpts initially perceived as speech and d' for excerpts initially perceived as song (Table 3). As predictors were correlated (Table 2), for each dependent variable we also conducted a multiple linear regression analysis to determine which of our four individual difference variables uniquely predicted variation in measures of the STS illusion.

Average trial-final ratings at session 1 were significantly positively correlated with Gold-MSI scores and tonal enculturation scores, but not with PEPS-C scores nor with attention to lyrics in music. Thus, participants with higher self-reported musical sophistication and higher performance on the tonal enculturation task gave higher average song-like ratings to all excerpts.

When all four predictors were entered into the regression model, only Gold-MSI scores uniquely predicted ratings. Therefore, initial ratings were primarily predicted by musical sophistication.

Number of excerpts heard as song at session 1 was also significantly positively correlated with Gold-MSI scores and tonal enculturation scores, but not with PEPS-C scores nor with attention to lyrics in music. This shows participants with higher self-reported musical sophistication and higher levels of tonal enculturation hear more excerpts as song-like, as was found for overall ratings. When all four predictors were entered into the regression model, only Gold-MSI scores uniquely predicted number of excerpts heard as song. This shows the number of excerpts that transform to song is primarily predicted by musical sophistication.

Transformation position at session 1 was significantly positively correlated with Gold-MSI scores and tonal enculturation scores, inversely correlated with PEPS-C scores, and shared no significant correlation with attention to lyrics. These correlations suggest participants with higher self-reported musical sophistication and higher levels of tonal enculturation require a greater number of repetitions to experience the STS illusion, as found in prior research (Falk et al., 2014). Also, participants who demonstrated a higher level of prosodic speech skills required less repetition to experience the STS illusion. When all four predictors were regressed onto transformation position, only PEPS-C scores significantly predicted transformation rate. Thus, for transformation position, prosodic skill uniquely predicted how quickly a participant perceived excerpts transform to song.

**Table 3.**

*Individual Difference Factors Predicting STS Illusion at Session 1*

|  | Average Ratings |  |  | Number of Excerpts Heard as Song |  |  | Transformation Position |  |  | Overall d' Score: Speech |  |  | Overall d' Score: Song |  |  |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Variable | r | β | SE | r | β | SE | r | β | SE | r | β | SE | r | β | SE |
| Gold-MSI Score | **0.30**$^{***}$ | **0.29**$^{***}$ | 0.05 | **0.31**$^{***}$ | **0.28**$^{***}$ | 0.38 | **0.15**$^{*}$ | 0.15 | 0.09 | 0.05 | 0.01 | 0.06 | **0.28**$^{***}$ | **0.26**$^{**}$ | 0.05 |
| Tonal Enculturation | **0.19**$^{**}$ | 0.12 | 0.004 | **0.20**$^{**}$ | 0.14 | 0.03 | **0.18**$^{*}$ | 0.14 | 0.01 | -0.02 | 0.001 | 0.004 | **0.17**$^{*}$ | 0.09 | 0.004 |
| PEPS-C (Prosody) | 0.07 | 0.02 | 0.01 | 0.07 | 0.03 | 0.06 | **-0.14**$^{*}$ | **-0.17**$^{*}$ | 0.02 | 0.05 | 0.08 | 0.01 | **0.17**$^{*}$ | 0.12 | 0.01 |
| Attention to Lyrics | -0.01 | -0.05 | 0.04 | 0.008 | -0.03 | 0.28 | -0.01 | -0.07 | 0.07 | 0.09 | 0.11 | 0.04 | -0.06 | -0.07 | 0.04 |
|  | $R^2 = 0.11$, $R^2$ = 0.09, $F(4, 153) = 4.83$, $p = .001$ |  |  | $R^2 = 0.34$, $R^2$ = 0.12, $F(4, 153) = 5.01$, $p < .001$ |  |  | $R^2 = 0.07$, $R^2$ = 0.05, $F(4, 153) = 2.98$, $p < .05$ |  |  | $R^2 = 0.01$, $R^2$ = -0.01, $F(4, 153) = 0.56$, $p > .05$ |  |  | $R^2 = 0.11$, $R^2$ = 0.09, $F(4, 153) = 4.88$, $p < .001$ |  |  |

r = simple correlation between predictor and outcome variable; β = standardized beta coefficients. * p < .05, ** p < .01, *** p < .001

Sensitivity (d’ score) for excerpts that did not transform to song (e.g., they were initially heard as speech at session 1) were not significantly correlated with any of our individual difference measures. This finding indicates none of our individual difference measures predicted discrimination performance for pitch changes when excerpts were perceived as speech. However, for excerpts that transformed to song at session 1, d’ scores were significantly correlated with Gold-MSI scores, tonal enculturation scores, and PEPS-C scores. Thus, participants with higher self-reported musical sophistication, higher levels of tonal enculturation, and better prosodic speech processing were better at detecting pitch changes to STS stimuli that were perceived as song. However, when all four predictors were entered into the regression model, only Gold-MSI scores were uniquely predictive.

Next, we asked whether our individual difference variables predicted participant rating consistency of STS excerpts across sessions. We used cross-session correlation across excerpts as our dependent measure (Table 4). Cross-session correlation averages were highly consistent for participants across sessions, indicating excerpts heard as song at session 1 are stably heard as song at session 2 (Table 1).

Cross-session correlations across excerpts were significantly negatively correlated with Gold-MSI scores and PEPS-C scores and positively correlated with tonal enculturation (see Table 4). Thus, participants with lower self-reported musical sophistication scores and lower levels of tonal enculturation showed higher cross-session correlations across excerpts. However, participants with higher levels of prosodic processing showed higher cross-session correlations. When all four predictors were entered into the regression model, Gold-MSI scores and PEPS-C scores uniquely predicted cross-session correlations. These results suggests that individuals who are more sensitive to speech prosody provided more consistent ratings for individual excerpts across sessions, whereas those with higher musical sophistication were less consistent; however, these groups were not mutually exclusive.

**Table 4.**

*Regression Results: Rating Consistency*

| | Cross-Session Correlations Across Excerpts | | |
|---|---|---|---|
| Variable | r | β | SE |
| Total Gold-MSI Score | **-0.24**$^{**}$ | **-0.23**$^{**}$ | 0.01 |
| Tonal Enculturation | **-0.15**$^{*}$ | -0.10 | 0.001 |
| PEPS-C (Prosody) | **0.15**$^{*}$ | **0.17**$^{*}$ | 0.002 |
| Attention to Lyrics | -0.12 | -0.05 | 0.01 |
| | $R^2 = 0.10$, $R^2_{\text{adj}} = 0.08$, $F(4, 153) = 4.28$, $p = .001$ | | |

r = simple correlation between predictor and outcome variable; β = standardized regression coefficients. * p < .05, ** p < .01, *** p < .001

# Chapter 4: Discussion

In the STS illusion, a listener experiences a perceptual shift in which an initial speech excerpt is heard to transform to song after several, unmodified repetitions (Deutsch et al., 2011). There exist many anecdotal reports that once a speech excerpt transforms to song, listeners continue to perceive it as song indefinitely, suggesting long-term stability of the STS illusion. However, one might also expect the STS illusion to decrease over delays because perceptual memories of previously heard STS stimuli may weaken over time. The current study empirically validated reports showing that when listeners hear excerpts transform to song at an initial session, those same excerpts are rated as more song-like than other excerpts at a second session up to 56 days later. Similarly, ratings of individual excerpts are correlated across sessions for the majority of listeners. These results provide strong support for the claim that once a given stimulus transforms to song for a listener, a stable, music-specific perceptual memory of that stimulus is formed (Dowling, 1991; Frieler et al., 2013; Levitin, 1994; Schellenberg & Habashi, 2015). Indeed, past research has shown listeners perceive word lists presented several times as opposed to once during an experimental session as more song-like, which also shows memory traces of increasingly familiar stimuli influence the short-term experience of the STS illusion (Soehlke et al., 2022).

Relatedly, Groenveld et al. (2020) showed that once a stimulus transforms to song, it remains heard as song even after fundamental frequency manipulations are made to its contour to make it sound increasingly less song-like. In their study, the contours of speech stimuli were manipulated to have decreasing, increasing, or random orders of $F_0$. Listeners who were first exposed to the condition for which the $F_0$ was maximized and then made to decrease across subsequent repetitions continued to provide higher overall ratings than for the increasing or random order conditions, thus demonstrating the influence of an initial memory trace formed for a song-like stimulus over short delays. These findings on perceptual memory in the auditory modality parallel work in the visual domain, which indicate the perception of a bistable stimulus may freeze on one representation due to the perceptual memory trace being formed between intermittent presentations of the stimulus (Leopold, Wilke, Maier, & Logothetis, 2002; Pearson & Brascamp, 2008). Many presentations of a stimulus can culminate into one representation which remains stable even when the stimulus is encountered at a later time, much like what occurs in the STS illusion where a percept of a stimulus as speech or song is formed over many repetitions of a speech excerpt. In line with these results in the visual domain, our results suggest the perceptual memory of an STS stimulus may be preserved for longer time frames of up to 56 days as shown by our cross-session correlations where participants provided similar ratings to individual stimuli across sessions (Pearson & Brascamp, 2008). However, the representation of an STS stimulus as speech- or song-like does not appear to be bistable—that is, once a listener hears an excerpt transform to song, they typically continue to hear any subsequent repetitions as song-like. Furthermore, if listeners assimilate tonal properties of transformed STS stimuli to the Western art scale at session 1, their formed musical schemas of those stimuli may enhance their perceptual memory, as representations of absolute musical properties like pitch and tempo have shown some stability across delays (Dowling, 1991; Durrant et al., 2015; Frieler et al., 2013; Levitin, 1994; Schellenberg & Habashi, 2015).

We also replicated previous findings that show, at initial exposure, semantic satiation of STS occurs between 3 to 4 repetitions of a speech excerpt, such that listeners no longer focus on the sematic content of the excerpt and instead focus on its musical characteristics such as beat-like patterns of syllables and prosodic speech cues (Castro, 2014; Castro et al., 2018; Falk et al., 2014; Vitevitch, et al., 2021). Additionally, session 1 song stimuli transformed to song fewer repetitions at session 2 (Figure 3), showing familiarity with STS stimuli more readily elicits a song-like percept upon future reencounters. Similar results were provided by Soehlke et al. (2022) where increased exposure to a speech stimulus elicited a stronger musical percept and higher song-like ratings at a later presentation within the same experimental session, especially when excerpts are repeated by the same versus a different speaker. Taken together, past and current results suggest the transformation of an STS stimulus from speech to song gives rise to a temporally stable musical percept which is quickly elicited upon future encounters.

Furthermore, in general, we found all stimuli were rated as more "song-like" at session 2, which resulted in the reclassification of several speech excerpts at session 1 as song at session 2. This result shows that repetition (i.e., increased exposure) affords greater song-like perception of STS stimuli across delays, regardless of whether they were initially perceived as speech or song. Repetition has been shown to elicit musical percepts of speech excerpts of various lengths, nonspeech sounds (Margulis & Simchy-Gross, 2016; Simchy-Gross & Margulis, 2018; Tierney et al., 2018), and speech signals decomposed into their rhythmic and spectral components, particularly with the use of shorter loop lengths between repetitions (Rowland, et al., 2019). To account for the general phenomenon of repetition increasing the perceived musicality of auditory stimuli, Rowland et al. (2019) coined the term "repetition to music effect," suggesting auditory input can be organized into preexisting internal musical representations based on the inherent qualities of the stimulus giving rise to the perception of "song," "music," "notes," or "beat" (Longuet-Higgins & Lee, 1982; Palmer & Krumhansl, 1990). The mechanisms which explain the repetition to music effect have yet to be established, however, Rowland et al. (2019) suggest the use of shorter loops lengths for auditory stimuli relies on earlier temporal windows of auditory processing which may permit speech and environmental sounds to be temporarily categorized outside of their typical, respective representations and be interpreted as musical. This theory supports a general mechanism for auditory processing that is biased toward increasing the perceived musicality of all forms of auditory input by means of repetition, which has been previously described as a "musical mode" of listening (Falk, 2014; Margulis, 2014).

At both sessions, we expected to conceptually replicate prior findings that detection of pitch changes was worse when those changes conformed to rather than violated the presumed musical scale and key of transformed excerpts, suggesting that listeners recruit music-specific cognitive representations of musical structure (Vanden Bosch der Nederlanden, et al., 2015b). Prior work also showed that when STS stimuli did not transform (i.e., were heard as speech), both conforming and non-conforming pitch changes were equally detectable, presumably because no music-specific expectations were activated. In the present study, listeners better detected non-conforming than conforming changes, regardless of whether the stimulus was initially perceived as speech or as song. Although the present findings partly replicated prior work, it is unclear why we found an advantage for detection of musical structure-violating changes regardless of initial percept. It is possible that task differences might have driven this difference, given that in Vanden Bosch der Nederlanden, et al.'s work (2015b), each trial started with the full context of the speech excerpt followed by 10 repetitions of an excerpt taken from context, consistent with the original task used in Deutsch et al. (2011). They also presented the detection task at the beginning *and* at the end of each trial, whereas in the current study, the discrimination task was performed in a separate block from the rating task.

Interestingly, results showed significant improvement across sessions in pitch discrimination only for stimuli categorized within the song domain. This improvement might reflect an effect of repetition enhancing pitch and tempo encoding for song-like stimuli and consolidating memory traces over time, such that when participants were presented with the same song-like excerpts at session 2, any violation to their stimulus-specific musical schema was increasingly aberrant and therefore readily detectable (Dowling, 1991; Durrant et al., 2015; Schellenberg & Habashi, 2015).

Superior discrimination for non-conforming pitch changes made to musical stimuli might be a function of the delay between initial and final percept of a stimulus, such that time elapsed between presentations allows for the increase in STS illusion stability—even for excerpts which did not transform to song. The short amount of time elapsed between the initial and final discrimination task in Vanden Bosch der Nederlanden, et al. (2015b) may have served to distinguish speech and song stimuli more markedly, as speech excerpts did not have as much time to begin transforming to song; therefore, discrimination for both conforming and non-conforming pitch changes made to speech stimuli were equally detectable. However, as evidenced by the current study, the STS illusion increased over delays even for stimuli initially heard as speech, perhaps making non-conforming pitch changes more detectable for both speech and song stimuli (Grenoveld et al., 2020; Pearson & Brascamp, 2008). Further research should be conducted using both methods to determine how significant a role increased presentations of a stimulus has on pitch change discrimination ability, while taking the amount of delay between initial and final percept of a stimulus into account.

For our individual difference analyses, we first examined correlational relationships between our individual difference measures. We found that Total Gold-MSI (musical sophistication) scores are significantly positively correlated with tonal enculturation scores and attention to lyrics. The significant correlation between musical sophistication and tonal enculturation scores indicates consistency between high scores on our self-reported musical sophistication and performance on our objective measure of tonal enculturation. Musical sophistication scores were also significantly positively correlated with attention to lyrics, which indicates participants that received a high musical sophistication score self-reported they generally pay attention to lyrical content while listening to music. This finding suggests participants' self-perception of musicality entails that they listen more closely to song lyrics, which might be reflected in their score on the Gold-MSI subtask dedicated to singing ability, as it calls for paying attention to lyrics. However, this finding could be considered counterintuitive, since it suggests participants who are more musical attend less to semantic—as opposed to prosodic—content of speech excerpts despite prosody being considered the "melody of language" (Boutsen, 2003). As such, gathering data on participants' attention to lyrical content from more objective measures would be useful in validating their self-report intuitions reflected by Gold-MSI scores. Conversely, individuals that indicated they pay more attention to lyrics shared a significantly negative correlational relationship with tonal enculturation. This finding is paradoxical because musical sophistication and tonal enculturation scores are significantly positively correlated measures assessing musicality, so it would seem that since attention to lyrics shares a significantly positively correlated with self-reported musical sophistication, it would share a similar relationship with tonal enculturation scores. Additionally, attention to lyrics shares a significantly negative correlational relationship with PEPS-C scores, such that those who self-reported they listen to lyrics received a lower score for prosodic sensitivity, presumably because attending to lyrics demands a greater focus on the semantic content of language and lesser focus on its more musical and prosodic elements (Bidelman & Lee, 2015; Jaisin et al., 2016).

We also examined how individual differences in musical sophistication (total Gold-MSI index score), sensitivity to speech prosody, tonal enculturation, and attention to lyrics predict the elicitation of the STS illusion at session 1. STS illusion elicitation was measured by mean ratings, number of excerpts heard as song, total d’ for speech stimuli, and total d’ for song stimuli. First, average ratings of stimuli were predicted by musical sophistication. This finding validates past research by Vanden Bosch der Nederlanden et al. (2015a) which shows that while both musicians and non-musicians experience the STS illusion (Mullin, 2021), musicians tend to provide higher ratings for STS stimuli. Musicians may have heard STS stimuli as more song-like due to their relying on both implicit and explicit memories to form robust musical representations, whereas non-musicians rely mostly on implicit knowledge (Cohen, et al., 2011; Tillmann, 2005).

Moreover, the number of stimuli that transformed to song was also predicted by musical sophistication. This result may reflect the tendency of those with high self-reported levels of musical sophistication to rate stimuli as more song-like on average. Our results are consistent with previous findings suggesting that while STS illusion strength might be linked to implicit musical sophistication measures, it is not linked to more objective measures such as performance on our tonal enculturation task. Previous research validates this finding, wherein years of musical training were not significantly correlated with performance on musical tasks or STS illusion elicitation (Tierney, Patel, Jasmin, & Breen, 2021). Instead, increased musical sophistication seems to increase general song-like perception, serving to elicit the STS illusion more frequently.

Transformation position at session 1 was an additional variable which measured the rate at which an excerpt transforms to song, where earlier transformation indicates faster transformation. As predicted, scores on the test of prosody (PEPS-C) inversely predicted transformation position at session 1. This result suggests that the less prosodic sensitivity exemplified by a participant on the PEPS-C, the later excerpts transform to song, validating prior evidence that shows focusing less on lexical content and more on musical characteristics of a speech excerpt (e.g., beat-like patterns and prosodic speech cues) supports more ready elicitation of the STS illusion (Castro, 2014; Castro et al., 2018; Falk et al., 2014, Margulis, 2014; Vitevitch, et al., 2020).

When total d’ score for speech was regressed onto our independent measures, none significantly predicted discrimination performance for both and conforming and non-conforming pitch changes; however, total d’ score for song was significantly predicted by musical sophistication scores for both types of pitch change. This finding indicates that both implicit and explicit musical knowledge may aid pitch discrimination performance for STS stimuli perceived as song-like (Cohen, et al., 2011; Tillmann, 2005). Furthermore, these results validate past predictions that participants with higher levels of musical sophistication might show better discrimination task performance than non-musicians more generally, regardless of delay, due to their increased sensitivity to pitch and enhanced cognitive control during working memory tasks (Besson, et al., 2007; Dowling et al., 2008; Schön et al., 2004).

Next, we examined rating consistency across sessions based on our individual difference measures. To measure rating consistency, we used cross-session correlations between session 1 and session 2 ratings since results showed ratings were significantly consistent over delays. Cross-session correlations were significantly inversely predicted by musical sophistication scores. This finding indicates that participants who self-reported higher levels of musical sophistication demonstrated less consistent ratings for individual stimuli across sessions, potentially because more STS excerpts initially rated as speech at session 1 transformed to song at session 2 (Table 4). Additionally, cross-session correlations were significantly positively predicted by PEPS-C scores, showing participants who have more prosodic sensitivity gave more consistent ratings for individual excerpts across sessions. This finding suggests prosodic sensitivity plays an important role in forming stable perceptual memories of STS stimuli as speech- or song-like.

Overall, we find strong support that the STS illusion is not only stable, but may increase over time, even across delays of up to 56 days. Moreover, we find support for the notion that the STS illusion is a general phenomenon of the human auditory system that is modulated by both implicit and explicit measures of musicality (Cohen et al., 2011; Tillmann, 2005). As shown in prior research (Mullin, 2021; Tierney et al., 2021; Vanden Bosch der Nederlanden, et al., 2015a), implicit knowledge of musical structure gained through everyday musical experience is sufficient to perceive the STS illusion, regardless of musical training.

Future research should investigate the STS illusion across development to determine how music- and language-specific knowledge is acquired and differentiated according to contextual demands and individual differences of musical and linguistic processing. This work offers additional information on how the auditory system interprets musical and linguistic stimuli and thus holds implications for the advancement of effective hearing devices (e.g., cochlear implants, hearing aids) and diagnostic assessments for individuals with communicative deficits. The results of this study contribute to basic science on the auditory system, as well as the stability of auditory memories while factoring in effects of individual differences.

# Appendix A

Compliance check questions and response options:

1) "What color is the sky? Please answer this incorrectly, on purpose, by choosing YELLOW instead of blue."

   Response options were Green, Red, Blue, or Yellow. Any participant who did not answer "Yellow" was excluded.

2) "People are working on this task in many different places. Please tell us about the place you worked on this task. Please answer honestly."

   Response options were: I worked on this study in a very noisy place, I worked on this study in a somewhat noisy place, I worked on this study in a somewhat quiet place, or I worked on this study in a very quiet place. Any participant who answered, "I worked on this study in a very noisy place" or "I worked on this study in a somewhat noisy place" and failed the headphone check was excluded.

3) "Did you wear headphones while listening to the sounds in this experiment? Please answer honestly." Response options were: Yes or No. Any participant who answered "No" was excluded.

4) "Students are working on this task with many different devices, browsers, and internet connections. Please tell us about whether you had difficulty loading the sounds. Please answer honestly." Response options were: There were problems loading all of the sounds, There were problems loading most of the sounds, There were problems loading some of the sounds, There were no problems loading any of the sounds. Any participant who answered with any response other than, "There were no problems loading any of the sounds" was excluded.

5) “How carefully did you complete this experiment? Please answer honestly.”

   Response options were: Not at all carefully, Slightly carefully, Moderately carefully, Quite carefully, or Very carefully. Any participant who answered “Not at all carefully” or “Slightly carefully” was excluded.

# References

Besson, M., Chobert, J., & Marie, C. (2011). Transfer of training between music and speech: common processing, attention, and memory. *Frontiers in psychology, 2*, 94.

Besson, M., Schön, D., Moreno, S., Santos, A., & Magne, C. (2007). Influence of musical expertise and musical training on pitch processing in music and language. *Restorative neurology and neuroscience, 25*(3-4), 399-410.

Bidelman, G. M., & Lee, C. C. (2015). Effects of language experience and stimulus context on the neural organization and categorical perception of speech. *Neuroimage*, *120*, 191-200.

Boersma, P., & Weenink, D. (2001). Praat, a system for doing phonetic by computer.

Boutsen, F. (2003). Prosody: the music of language and speech. *The ASHA leader*, *8*(4), 6-8.

Castro, N. (2014). *What Turns Speech into Song? Investigations of the Speech-to-Song Illusion* (Doctoral dissertation, University of Kansas).

Castro, N., Mendoza, J. M., Tampke, E. C., & Vitevitch, M. S. (2018). An account of the speech-to-song illusion using node structure theory. *PloS one*, 13(6).

Chen, X., Affourtit, J., Ryskin, R., Regev, T. I., Norman-Haignere, S., Jouravlev, O., ... & Fedorenko, E. (2021). The human language system does not support music processing. *BioRxiv*.

Chomsky, N. (1959). A review of B. F. Skinner's verbal behavior. *Language*, 35(1), 26–58.

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.

Cohen, M. A., Horowitz, T. S., & Wolfe, J. M. (2009). Auditory recognition memory is inferior to visual recognition memory. *Proceedings of the National Academy of Sciences*, *106*(14), 6008-6010.

Cosmides, L., & Tooby, J. (1994). *Origins of domain specificity: The evolution of functional organization. Mapping the mind: Domain specificity in cognition and culture.* Cambridge, UK: Cambridge University Press, pp. 85–116.

Cutler, A. (1991). Linguistic rhythm and speech segmentation. In *Music, language, speech and brain* (pp. 157-166). Palgrave, London.

Dalla Bella, S., Peretz, I., & Aronoff, N. (2003). Time course of melody recognition: A gating paradigm study. *Perception & Psychophysics, 65*(7), 1019-1028.

Dehaene-Lambertz, G., Pallier, C., Serniclaes, W., Sprenger-Charolles, L., Jobert, A., & Dehaene, S. (2005). Neural correlates of switching from auditory to speech perception. *Neuroimage*, *24*(1), 21-33.

Deutsch, D. (2003). But they sometimes behave so strangely. *Phantom Words and Other Curiosities.*

Deutsch, D., Henthorn, T., & Lapidis, R. (2011). The illusory transformation of speech to song. *Journal of the Acoustical Society of America*, 129(4), 2245–2252.

Dibben, N., Coutinho, E., Vilar, J. A., & Estévez-Pérez, G. (2018). Do individual differences influence moment-by-moment reports of emotion perceived in music and speech prosody?. *Frontiers in Behavioral Neuroscience*, *12*, 184.

Dowling, W. J. (1991). Tonal strength and melody recognition after long and short delays. *Perception & Psychophysics*, *50*(4), 305-313.

Dowling, W. J., Bartlett, J. C., Halpern, A. R., & Andrews, M. W. (2008). Melody recognition at fast and slow tempos: Effects of age, experience, and familiarity. *Perception & psychophysics*, *70*(3), 496-502.

Durrant, S. J., Cairney, S. A., McDermott, C., & Lewis, P. A. (2015). Schema-conformant memories are preferentially consolidated during REM sleep. *Neurobiology of Learning and Memory, 122*, 41-50.

Falk, S., & Rathcke, T. (2010). On the speech-to-song illusion: Evidence from German. In *Speech Prosody 2010-Fifth International Conference*.

Falk, S., Rathcke, T., & Dalla Bella, S. (2014). When speech sounds like music. *Journal of Experimental Psychology: Human Perception and Performance*, 40(4), 1491.

Fedorenko, E., McDermott, J. H., Norman-Haignere, S., & Kanwisher, N. (2012). Sensitivity to musical structure in the human brain. *Journal of neurophysiology, 108*(12), 3289-3300.

Fodor, J. A. (1983). *The modularity of mind: An essay on faculty psychology*. MIT Press.

Frieler, K., Fischinger, T., Schlemmer, K., Lothwesen, K., Jakubowski, K., & Müllensiefen, D. (2013). Absolute memory for pitch: A comparative replication of Levitin's 1994 study in six European labs. *Musicae Scientiae, 17*(3), 334-349.

Fujioka, T., Trainor, L. J., Ross, B., Kakigi, R., & Pantev, C. (2005). Automatic encoding of polyphonic melodies in musicians and nonmusicians. *Journal of cognitive neuroscience*, *17*(10), 1578-1592.

Graber, E., Simchy-Gross, R., & Margulis, E. H. (2017). Musical and linguistic listening modes in the speech-to-song illusion bias timing perception and absolute pitch memory. *The Journal of the Acoustical Society of America, 142*(6), 3593-3602.

Groenveld, G., Burgoyne, J. A., & Sadakata, M. (2020). I still hear a melody: investigating temporal dynamics of the Speech-to-Song Illusion. *Psychological Research, 84*(5), 1451-1459.

Guo, S., & Koelsch, S. (2016). Effects of veridical expectations on syntax processing in music: Event-related potential evidence. *Scientific reports*, *6*(1), 1-11.

Hansen, N. C., Vuust, P., & Pearce, M. (2016). " If You Have to Ask, You'll Never Know": Effects of Specialised Stylistic Expertise on Predictive Processing of Music. *PLoS One, 11*(10), e0163584.

Hautus, M. J. (1995). Corrections for extreme proportions and their biasing effects on estimated values ofd′. *Behavior Research Methods, Instruments, & Computers*, *27*(1), 46-51.

Hébert, S., Racette, A., Gagnon, L., & Peretz, I. (2003). Revisiting the dissociation between singing and speaking in expressive aphasia. *Brain*, *126*(8), 1838-1850.

Jackendoff, R. (2009). Parallels and nonparallels between language and music. *Music perception, 26(3)*, 195-204.

Jaisin, K., Suphanchaimat, R., Figueroa Candia, M. A., & Warren, J. D. (2016). The speech-to-song illusion is reduced in speakers of tonal (vs. non-tonal) languages. *Frontiers in psychology*, 7, 662.

Joanisse, M. F., & Gati, J. S. (2003). Overlapping neural regions for processing rapid temporal cues in speech and nonspeech signals☆. *Neuroimage*, *19*(1), 64-79.

Karmiloff-Smith, A. (2009). Nativism versus neuroconstructivism: rethinking the study of developmental disorders. *Developmental psychology, 45*(1), 56.

Kauffman, W. H., & Carlsen, J. C. (1989). Memory for intact music works: The importance of music expertise and retention interval. *Psychomusicology: A Journal of Research in Music Cognition*, *8*(1), 3.

Leopold, D. A., Wilke, M., Maier, A., & Logothetis, N. K. (2002). Stable perception of visually ambiguous patterns. *Nature neuroscience*, *5*(6), 605-609.

Leung, C., & Zhou, D. H. R. (2018). Pace, Emotion, and Language Tonality on Speech-to-song Illusion.

Levitin, D. J. (1994). Absolute memory for musical pitch: Evidence from the production of learned melodies. *Perception & Psychophysics, 56*(4), 414-423.

Liu, F., Patel, A. D., Fourcin, A., & Stewart, L. (2010). Intonation processing in congenital amusia: discrimination, identification and imitation. *Brain, 133*(6), 1682-1693.

Longuet-Higgins, H. C., & Lee, C. S. (1982). The perception of musical rhythms. *Perception*, *11*(2), 115-128.

Lu, X., Ho, H. T., Liu, F., Wu, D., & Thompson, W. F. (2015). Intonation processing deficits of emotional words among Mandarin Chinese speakers with congenital amusia: an ERP study. *Frontiers in Psychology, 6*, 385.

Luo, H., & Poeppel, D. (2012). Cortical oscillations in auditory perception and speech: evidence for two temporal windows in human auditory cortex. *Frontiers in psychology, 3*, 170.

MacKay, D. G. (1987*). The organization of perception and action: A theory for language and other cognitive skills.* Springer-Verlag.

Magne, C., Schön, D., & Besson, M. (2006). Musician children detect pitch violations in both music and language better than nonmusician children: behavioral and electrophysiological approaches. *Journal of cognitive neuroscience, 18*(2), 199-211.

Margulis, E. H. (2014*). On repeat: How music plays the mind*. Oxford University Press.

Margulis, E. H., & Simchy-Gross, R. (2016). Repetition enhances the musicality of randomly generated tone sequences. *Music Perception: An Interdisciplinary Journal*, *33*(4), 509-514.

Margulis, E. H., Simchy-Gross, R., & Black, J. L. (2015). Pronunciation difficulty, temporal regularity, and the speech-to-song illusion. *Frontiers in Psychology*, 6, 1–7.

Marmel, F., & Tillmann, B. (2009). Tonal priming beyond tonics. *Music Perception, 26*(3), 211-221.

Marmel, F., Tillmann, B., & Dowling, W. J. (2008). Tonal expectations influence pitch perception. *Perception & Psychophysics, 70*(5), 841-852.

McAuley, J. D., Stevens, C., & Humphreys, M. S. (2004). Play it again: Did this melody occur more frequently or was it heard more recently? The role of stimulus familiarity in episodic recognition of music. *Acta Psychologica, 116*(1), 93-108.

Mehr, S. A., Singh, M., York, H., Glowacki, L., & Krasnow, M. M. (2018). Form and function in human song. Current Biology, 28(3), 356-368.

Möttönen, R., Calvert, G. A., Jääskeläinen, I. P., Matthews, P. M., Thesen, T., Tuomainen, J., & Sams, M. (2006). Perceiving identical sounds as speech or non-speech modulates activity in the left posterior superior temporal sulcus. *Neuroimage, 30*(2), 563-569.

Müllensiefen, D., Gingras, B., Musil, J., & Stewart, L. (2014). The musicality of non-musicians: An index for assessing musical sophistication in the general population. *PloS one, 9*(2), e89642.

Mullin, H. A., Norkey, E. A., Kodwani, A., Vitevitch, M. S., & Castro, N. (2021). Does age affect perception of the Speech-to-Song Illusion?. *Plos one*, *16*(4), e0250042.

Norman-Haignere, S., Kanwisher, N. G., & McDermott, J. H. (2015). Distinct cortical pathways for music and speech revealed by hypothesis-free voxel decomposition. *Neuron,* 88(6), 1281-1296.

Pallesen, K. J., Brattico, E., Bailey, C. J., Korvenoja, A., Koivisto, J., Gjedde, A., & Carlson, S. (2010). Cognitive control in auditory working memory is enhanced in musicians. *PloS one, 5*(6), e11120.

Palmer, C., & Krumhansl, C. L. (1990). Mental representations for musical meter. *Journal of Experimental Psychology: Human Perception and Performance*, *16*(4), 728.

Park, J. M., Chung, C. K., Kim, J. S., Lee, K. M., Seol, J., & Yi, S. W. (2018). Musical expectations enhance auditory cortical processing in musicians: A magnetoencephalography study. *Neuroscience, 369*, 325-335.

Patel, A. D., Foxton, J. M., & Griffiths, T. D. (2005). Musically tone-deaf individuals have difficulty discriminating intonation contours extracted from speech. *Brain and cognition, 59*(3), 310-313.

Pearson, J., & Brascamp, J. (2008). Sensory memory for ambiguous vision. *Trends in cognitive sciences*, *12*(9), 334-341.

Peppé, S., & McCann, J. (2003). Assessing intonation and prosody in children with atypical language development: the PEPS-C test and the revised version. *Clinical Linguistics & Phonetics*, *17*(4-5), 345-354.

Peretz, I. (2001). Brain specialization for music: New evidence from congenital amusia. *Annals of the New York Academy of Sciences*, *930*(1), 153-165.

Peretz, I., & Hyde, K. L. (2003). What is specific to music processing? Insights from congenital amusia. *Trends in cognitive sciences*, *7*(8), 362-367.

Pinheiro, A. P., Vasconcelos, M., Dias, M., Arrais, N., & Gonçalves, Ó. F. (2015). The music of language: An ERP investigation of the effects of musical training on emotional prosody processing. *Brain and Language, 140*, 24-34.

Poeppel, D. (2003). The analysis of speech in different temporal integration windows: cerebral lateralization as ‘asymmetric sampling in time’. *Speech communication*, *41*(1), 245-255.

Remez, R., Rubin, P., Pisoni, D., & Carrell, T. (1981). Speech perception without traditional speech cues. *Science*, 212(4497), 947–949.

Rogalsky, C., Rong, F., Saberi, K., & Hickok, G. (2011). Functional anatomy of language and music perception: temporal and structural factors investigated using functional magnetic resonance imaging. *Journal of Neuroscience*, *31*(10), 3843-3852.

Rowland, J., Kasdan, A., & Poeppel, D. (2019). There is music in repetition: Looped segments of speech and nonspeech induce the perception of music in a time-dependent manner. *Psychonomic Bulletin & Review*, *26*(2), 583-590.

Schellenberg, E. G., & Habashi, P. (2015). Remembering the melody and timbre, forgetting the key and tempo. *Memory & Cognition*, *43*(7), 1021-1031.

Schön, D., Magne, C., & Besson, M. (2004). The music of speech: Music training facilitates pitch processing in both music and language. *Psychophysiology*, *41*(3), 341-349.

Severance, E., & Washburn, M. F. (1907). The loss of associative power in words after long fixation. *The American Journal of Psychology*, 182-186.

Simchy-Gross, R., & Margulis, E. H. (2018). The sound-to-music illusion: Repetition can musicalize nonspeech sounds. *Music & Science*, *1*, 2059204317731992.

Soehlke, L. E., Kamat, A., Castro, N., & Vitevitch, M. S. (2022). The influence of memory on the speech-to-song illusion. *Memory & Cognition*, 1-12

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior research methods, instruments, & computers*, *31*(1), 137-149.

Tervaniemi, M., & Hugdahl, K. (2003). Lateralization of auditory-cortex functions. *Brain research reviews*, *43*(3), 231-246.

Thompson, W. F., Marin, M. M., & Stewart, L. (2012). Reduced sensitivity to emotional prosody in congenital amusia rekindles the musical protolanguage hypothesis. *Proceedings of the National Academy of Sciences*, 109(46), 19027–19032.

Tierney, A., Dick, F., Deutsch, D., & Sereno, M. (2013). Speech versus song: Multiple pitch-sensitive areas revealed by a naturally occurring musical illusion. *Cerebral Cortex*, 23, 249–254.

Tierney, A., Patel, A. D., & Breen, M. (2018). Repetition enhances the musicality of speech and tone stimuli to similar degrees. *Music Perception: An Interdisciplinary Journal*, 35(5), 573-578.

Tierney, A., Patel, A. D., Jasmin, K., & Breen, M. (2021). 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.

Tierney, A., Patel, A., & Breen, M. (2016). Acoustic and musical foundations of the speech/song illusion. In *Proceedings of the 14th International Conference on Music Perception and Cognition* (pp. 369-373).

Tillmann, B. (2005). Implicit investigations of tonal knowledge in nonmusician listeners. *Annals of the New York Academy of Sciences*, *1060*(1), 100-110.

Trainor, L. J., & Trehub, S. E. (1994). Key membership and implied harmony in Western tonal music: Developmental perspectives. *Perception & Psychophysics*, *56*(2), 125-132.

Vanden Bosch der Nederlanden, C. M.., Hannon, E. E., & Snyder, J. S. (2015b). Finding the music of speech: Musical knowledge influences pitch processing in speech. *Cognition*, *143*, 135-140.

Vanden Bosch der Nederlanden, C.M., Hannon, E. E., & Snyder, J. S. (2015a). Everyday musical experience is sufficient to perceive the speech-to-song illusion. *Journal of experimental psychology: General*, *144*(2), e43.

Vitevitch, M. S., Ng, J. W., Hatley, E., & Castro, N. (2021). Phonological but not semantic influences on the speech-to-song illusion. *Quarterly Journal of Experimental Psychology*, *74*(4), 585-597.

Woods, K., Siegel, M. H., Traer, J., & McDermott, J. H. (2017). Headphone screening to facilitate web-based auditory experiments. *Attention, perception & psychophysics*, *79*(7), 2064–2072. https://doi.org/10.3758/s13414-017-1361-2

Zatorre, R. J., & Gandour, J. T. (2008). Neural specializations for speech and pitch: moving beyond the dichotomies. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1493), 1087-1104.

Zatorre, R. J., Belin, P., & Penhune, V. B. (2002). Structure and function of auditory cortex: Music and speech. *Trends in Cognitive Sciences*, 6(1), 37–46.

Zatorre, R., Evans, A., Meyer, E., & Gjedde, A. (1992). Lateralization of phonetic and pitch discrimination in speech processing. *Science,* 256(5058), 846–849.

Zuk, J., Benjamin, C., Kenyon, A., & Gaab, N. (2014). Behavioral and neural correlates of executive functioning in musicians and non-musicians. *PloS one*, *9*(6), e99868.

# Curriculum Vitae

# Rodica R. Constantine

rodi.constantine@gmail.com **Citizenship: United States**

---

## EDUCATION

- **Ph.D. Candidate, Developmental Emphasis** *Fall 2018 -present*
  Psychological and Brain Sciences, *University of Nevada, Las Vegas*
  Graduate Student, Auditory Cognition and Development Lab Advisor:
  Erin E. Hannon, Ph.D.

- **Master of Arts (Psychology)** *Fall 2018 - Fall 2022*
  Psychology, *University of Nevada, Las Vegas*

  Title: The Long-term Stability of the Speech-to-Song-Illusion and Individual Differences Advisor:
  Erin E. Hannon, Ph.D.

- **Bachelor of Science (B.S. Magna cum Laude, Honors Certificate)** *Fall 2012 - Spring2018*
  Psychology, *University of Pittsburgh*

- **Bachelor of Arts (B.A. Magna cum Laude, Honors Certificate)** *Fall 2012 - Spring2018*
  Music, *University of Pittsburgh*
  Piano Instructor: Luz Manriquez

- **Graduate Certificate** *Fall 2015 - Spring2018*
  Piano Pedagogy, *Carnegie Mellon University*
  Supervisors: Hanna Wu Li, Luz Manriquez

## RESEARCH INTERESTS

- Developmental Cognitive Neuroscience, Auditory Cognition, Music and Language Perception, Auditory Scene Analysis, Affective Auditory Processing, Linguistics, Intersensory Perception

## LANGUAGES

- English (Native), Romanian (Native), German (A2), Italian(A2)

## PROGRAMMING LANGUAGES AND SCIENTIFIC SOFTWARE

- Audacity,Excel,GoldWave,JASP,LaTex,MatLab,OpenSesame,PRAAT,Psychtoolbox,Presentation,Qualtrics, R, Splus, SPSS

## PUBLICATIONS

- **Constantine, R. R.**, Getty, D. J., & Fraundorf, S. H. (2022). The Role of Priming in Grammatical Acceptability Judgements for Native versus Non-Native Speakers: Effects of Intelligibility. *Plos one*, *17*(9), e0275191.

- Constantine, G. M., Buliga, M. G., **Constantine, R. R.** , Md, R. A. N., & Md, T. R. B. (2022). Rank-Based Models for Predicting Trauma-Related Organ Failure. *Biomedical Journal of Scientific and Technical Research, 44*(2), 35283-35287.

## CONFERENCE PRESENTATIONS

- **Constantine, R.R.** & Hannon, E. E. (2023, February). *The Long-Term Stability of the Speech-to-Song Illusion.* Poster Presentation at Association for Research in Otolaryngology (ARO), Orlando, FL.

- **Constantine,R.R.** & Hannon, E. E. (2022, November). *The Long-Term-Stability of the Speech-to-Song Illusion and Individual Differences.* Presentation at Auditory Perception, Cognition, and Action Meeting (APCAM), Boston, MA.
  **Also presented at...**
  – Psychonomic Society, 63rd Annual Meeting, Boston, MA. (2022, November).
  – Society for Music Perception and Cognition (SMPC), Portland, OR. (2022, August).

- Barashy, S., Mednicoff, S. D., **Constantine, R. A.**, Benning, S. D., Snyder, J.S., and Hannon, E. E. (2022, November). *Misophonic Experienceand Musicality.* Poster Presentation at Auditory Perception, Cognition, and Action Meeting (APCAM), Boston, MA.
  **Also presented at...**
  – Association for Research in Otolaryngology (ARO), Orlando, FL. (2023, February).

- **Constantine, R.R.**& Fraudorf, S. H. (2018, November). *Syntactic Adaptation Between Nativeand Non-Native Speakers.* Poster Presentation at Psychonomic Society, 59th Annual Meeting, New Orleans, LA.

## TALKS

- **Constantine, R.R.** (2023, April).*The Long-Term Stability of the Speech-to-Song Illusion.* Presentation at the Graduate & Professional Student Research Forum. Las Vegas, NV.
- **Constantine, R.R.** (2022, December).*The Long-Term Stability of the Speech-to-Song Illusion and Individual Differences.* Master's Thesis Defense, Department of Psychology, University of Nevada, Las Vegas, NV.
- **Constantine, R.R.** (2022, April).*The Long-Term Stability of the Speech-to-Song Illusion and Individual Differences.* Presentation at the UNLV Psychology Department Research Fair. Las Vegas, NV.
- **Constantine,R.R.** (2020, January). *The Speech-to-Song Illusion Across Development.* Presentation at the UNLV Psychology Department Proseminar Course. Las Vegas, NV.
- **Constantine, R.R.** (2019, April).*The Speech-to-Song Illusion Across Development.* Presentation at the UNLV Psychology Department Proseminar Course. Las Vegas, NV.

## ONGOING RESEARCH PROJECTS

- **The Long-Term Stability of the Speech-to-Song Illusion**
- **The Speech-to-Song Illusion and Individual Differences**
- **Perfect Pitch and the Speech-to-Song Illusion**
- **The Speech-to-Song Illusion: An Acoustic Analysis**
  – An acoustic analysis of Speech-to-Song Illusion stimuli
- **The Speech-to-Song Illusion Across Development**
  – Investigating the Speech-to-Song Illusion across various developmental age groups to gauge how and when differences in music- and language- specific processing emerge
- **The Auditory Continuity Illusion**
  – Exploring the effects of familiarity on the neural correlates of illusory music perception
- **Perceptual Learning Across Domains (PLaD)**
  – Researching perceptual learning across the domains of music and language within the first year of development.
  – Created multiple programs in R to conduct all inter-rater reliability analyses
- **ManyBabies 3 (MB3)**
  – Examining the statistical learning in language for infants 5 to 10 months old as part of a large-scale multi-lab collaboration

## RESEARCH EXPERIENCE

- **Lab Manager**, Auditory Cognition and Development Lab *Fall 2018 - Present*
  *University of Nevada, Las Vegas*
- Memory and Psycho-linguistics in Learning & Education (MAPLE) Lab *Fall 2017 – Spring 2018*
  PI: Scott H. Fraundorf, Ph.D.
  *Research Assistant, University of Pittsburgh*
- Learning Imaging & Family Experience (LIFE) Lab *Fall 2017 – Spring 2018*
  PI: Jamie L. Hanson, Ph.D.
  *Research Assistant, University of Pittsburgh*

## FELLOWSHIPS, SCHOLARSHIPS, CERTIFICATES, and AWARDS

- **Patricia Sastaunik Scholarship** *Fall 2021 – Spring 2023*
  *University of Nevada, Las Vegas*
- **Graduate & Professional Student Association Travel Grant** *Summer 2022*
  *University of Nevada, Las Vegas*
- **Summer DoctoralResearch Fellowship** *Summer 2022*
  *University of Nevada, Las Vegas*
- **Inaugural Excellence in Teaching and Mentoring Award** *Fall 2021*
  *University of Nevada, Las Vegas*
- **Online Teaching EssentialsCourse Certificate** *Fall 2020*
  *University of Nevada, Las Vegas*
- **Summer Session Scholarship** *Summer 2020, Summer2021*
  *University of Nevada, Las Vegas*
- **UNLV Access Grant** *Fall 2018*
  *University of Nevada, Las Vegas*

- **Anita J. Curka Music Scholarship** *Spring 2017 – Fall 2017*
  *University of Pittsburgh*

- **Semi-Finalist at the Cincinnati World Piano Competition** *Summer 2013*
  *Preliminary score: 98/100*

- **Selected as a pre-concert pianist at Heinz Hall** *Spring 2009 - 2012*
  *Steinway Society of Western PA*

## TEACHING EXPERIENCE

- **PSY 330, Foundations of Developmental Psychology: Infant and Child** *Fall 2020 - Present*
  Instructor, *University of Nevada, Las Vegas*
- **PSY 101, Introduction to Psychology** *Summer 2019*
  Instructor, *University of Nevada, Las Vegas*
- **Beginning Piano for Children II** *Summer 2017 - Fall2017*
  Instructor, *Carnegie Mellon University*
  Supervisors: Hanna Wu Li, Luz Manriquez
- **Beginning Piano for Children I** *Spring 2017*
  Instructor, *Carnegie Mellon University*
  Supervisors: Hanna Wu Li, Luz Manriquez

## Teaching Assistant

- **PSY 330, Foundations of Developmental Psychology: Infant and Child** *Fall 2018, 2019 - Spring 2020*
  Instructor: Erin Hannon, Ph.D.
  *University of Nevada, Las Vegas*
- **PSY 423, Language Development** *Spring 2019*
  Instructor: Erin Hannon, Ph.D.
  *University of Nevada, Las Vegas*

## PROFESSIONAL AFFILIATIONS

- **Psychonomic Society, Member** *Fall 2022 - present*
- **Association for Research in Otolaryngology (ARO), Member** *Fall 2022 - present*
- **Society for Music Perception and Cognition (SMPC), Member** *Spring 2019 - present*
- **Piano Teachers Connect, Instructor** *Fall 2019 - present*
- **Pittsburgh Concert Society, Patron** *Fall 2015 - present*

## PROFESSIONAL REFERENCES

**Dr. Erin Hannon**
*Associate Professor*, Department of Psychology
University of Nevada, Las Vegas
Phone: +1 702-895-4687
E-mail:erin.hannon@unlv.edu

**Dr. Joel Snyder**
*Associate Professor*, Department of Psychology
University of Nevada, Las Vegas
Phone: +1 702-895-4692
E-mail:joel.snyder@unlv.edu

**Dr. Diego Vega**
*Associate Professor*, Department of Music
University of Nevada, Las Vegas
Phone: +1 702-895-5040
E-mail:diego.vega@unlv.edu

**Dr. Scott Fraundorf**
*Associate Professor*, Department of Psychology
University of Pittsburgh
Phone: +1412-624-7029
E-mail:sfraundo@pitt.edu

**Dr. Richard Randall**
*Associate Professor*, Department of Music
Carnegie Mellon University
Phone: +1773-263-3620
E-mail:randall@cmu.edu