Finding the Music of Speech:
Musical Knowledge Influences Pitch Processing in Speech

Christina M. Vanden Bosch der Nederlanden^a, Erin E. Hannon^a, & Joel S. Snyder^a

^aUniversity of Nevada, Las Vegas

Word Count: 4,437

Corresponding Author:

C. M. Vanden Bosch der Nederlanden

University of Nevada, Las Vegas

4505 S. Maryland Pkwy

Las Vegas, NV 89154

USA

702-895-2995

vandenbo@unlv.nevada.edu

Abstract

Few studies comparing music and language processing have adequately controlled for low-level acoustical differences, making it unclear whether differences in music and language processing arise from domain-specific knowledge, acoustic characteristics, or both. We controlled acoustic characteristics by using the speech-to-song illusion, which often results in a perceptual transformation to song after several repetitions of an utterance. Participants performed a same-different pitch discrimination task for the initial repetition (heard as speech) and the final repetition (heard as song). Better detection was observed for pitch changes that violated rather than conformed to Western musical scale structure, but only when utterances transformed to song, indicating that music-specific pitch representations were activated and influenced perception. This shows that music-specific processes can be activated when an utterance is heard as song, suggesting that the high-level status of a stimulus as either language or music can be behaviorally dissociated from low-level acoustic factors.

Keywords: music, language, domain-specific, auditory illusion

Finding the Music of Speech: Musical Knowledge Influences Pitch Processing in Speech

1. Introduction

Music and language are rich and dynamic sound structures, intimately tied to human communication. These similarities have led to considerable speculation and debate over whether human listeners process musical and linguistic sounds using shared or distinct mechanisms. Central to this controversy are questions about the extent to which speech or music is “special” (Pinker & Jackendoff, 2005; Peretz & Hyde, 2003), what these unique capacities tell us about human nature and origins (Hauser, Chomsky, & Fitch, 2002), and whether transfer of skill can be observed between musical and linguistic abilities (Patel, 2011). Since the 1950’s, theorists have speculated that dedicated modules in the brain evolved through natural selection to detect and process music and language (Fodor, 1983; Chomsky, 1959; Cosmides & Tooby, 1994). Indeed, many studies are consistent with modular accounts, which predict that the status of the stimulus as either music or language drives the recruitment of encapsulated, domain-specific mechanisms. Musical and linguistic stimuli, such as chords and syllables, elicit distinct patterns of activation in the brain that tend to be right-lateralized for music and left-lateralized for language (e.g., Zatorre, Evans, Meyer, & Gjedde, 1992). Evidence for double dissociations of music and language processing comes from adults with congenital amusia, who have impaired pitch processing but perceive language normally (Peretz et al., 2002) and those with aphasia due to stroke, who have impaired language production but can nevertheless sing lyrics (see Merrett, Peretz, & Wilson, 2014).

By contrast, there is growing evidence for overlap of music and language processing, implicating domain-general mechanisms (Newport, 2011). For example, syntactic violations in language can interfere with syntactic processing of music and vice versa (Fedorenko et al.,

2009). Likewise, syntactic violations in both music and language have been shown to elicit the same patterns of activation in the brain (Patel et al., 1998; Levitin & Menon, 2003), although this could result from task-related similarities that place demands on working memory and/or executive function (Rogalsky, Rong, Saberi, Hickok, 2011). Claims of a pure double dissociation of music and language processing are also undermined by evidence that adults with congenital amusia have impaired processing of emotional speech prosody (Thompson, Marin & Stewart, 2012).

Even if music- and language-specific processes are observed among adults, this does not mean that such processes are innately domain-specific. Specialization could result from developmental processes and interaction with language and music throughout childhood (Karmiloff-Smith, 2009). Statistical learning mechanisms allow infants to learn the distributional and transitional probabilities in syllables, tones, and even visual shapes (Saffran, Aslin, & Newport, 1996; Saffran, Johnson, Aslin, & Newport, 1999; Kirkham, Slemmer, & Johnson, 2002). Infant listeners might broadly apply such mechanisms to interpret both spoken and sung input until they have acquired music- or language-specific knowledge or have learned to track specific acoustic cues that are critical within each domain (Zatorre & Gandour, 2008).

One problem shared by all investigations of music and language processing is that low-level acoustic characteristics may drive responses, regardless of the status of a stimulus as music or speech. For example, a musical tone with a fast sweep elicits greater left hemisphere activation whereas a vowel produces greater right hemisphere activation (Joanisse & Gati, 2003). Lateralized brain responses to such stimuli could arise entirely from their acoustic features, since the right hemisphere is presumed to be optimized for longer timescales, while the left hemisphere is optimized for extracting information from shorter timescales (Poeppel, 2003;

Tervaniemi & Hugdahl, 2003), or spectral vs. temporal processing for right and left hemispheres, respectively (Zatorre, Belin, & Penhune, 2002). Without controlling for differences in low-level acoustic characteristics of music and language stimuli, many of the studies claiming to index domain-specificity may in fact simply illustrate how the brain processes different acoustic characteristics.

It is unclear whether the acoustic characteristics of speech and song can be divorced from one's perception of a stimulus as music or language. If perception and interpretation of sound is entirely domain-general, then there should be no differences in behavioral or neural responses when acoustic characteristics of stimuli are held constant. If, on the other hand, the same acoustic characteristics give rise to different responses depending on the listener's percept of the stimulus as music or language, this would implicate domain-specific knowledge or processes. There is some indication that speech processing is not fundamentally tied to its acoustic characteristics, because sine-wave speech (SWS) perceived initially as noise or whistles can be heard as speech after training, a transition in perception that is accompanied by a change from right to left lateralization (Remez, Rubin, Pisoni, & Carrell, 1981; Dehaene-Lambertz et al., 2005). However, Dehaene-Lambertz and colleagues (2005) reported that SWS resulted in greater left-hemisphere activation compared to silence even when perceived as non-speech, suggesting greater activation observed for SWS perceived as speech could be due to increased attention or increased effort to identify sounds that are already processed preferentially by the left hemisphere. Furthermore, SWS studies do not allow for the comparison of linguistic and musical modes of processing, because when SWS is not heard as speech it is not heard as music either. The present paper will examine music-specific processing using a similar approach with a stimulus that can be perceived as speech or song.

Recently, researchers reported an auditory illusion in which a single natural speech excerpt repeated 10 times in succession was perceived by musicians and non-musicians to transform from speech to song (Deutsch, 2003; Deutsch, Henthorn, & Lapidis, 2011; Tierney, Dick, Deutsch, & Sereno, 2012; Vanden Bosch der Nederlanden, Hannon, & Snyder, 2015; Margulis, Simchy-Gross, & Black, 2015). When musicians were asked to vocally reproduce the stimulus after either one or ten repetitions, participants' reproductions after ten repetitions more closely matched the pitch contour of the spoken sentence but were also subtly altered to more closely fit the intervals of the Western musical scale, suggesting that repetition led musicians to process these excerpts through a Western musical lens (Deutsch et al., 2011). This result thus provides indirect evidence that the same physical stimulus might give rise to distinct percepts depending on whether the stimulus is heard as speech or music. However it is also possible the result was an artifact of using a vocal production task, leading sung reproductions to closely fit the musical scale because musicians are trained to sing that way. To more directly address the recruitment of musical knowledge, we developed a perceptual, non-vocal task to examine whether or not musicians and non-musicians recruit music-specific knowledge when they hear an excerpt as song.

Music-specific knowledge can help or hinder pitch perception. In Western music, melodies are typically composed of pitches that belong to a particular key or scale, whose members sound stable together relative to non-members. Consequently, in a melody discrimination task, experienced listeners more readily detect a pitch change that violates or goes outside of the key than a pitch change that conforms to the established key (Trainor & Trehub, 1992; 1994). We expanded this basic approach for use in the current study. We asked how well listeners detect pitch changes that conform to or violate the inferred Western musical scale

pitches of speech-to-song excerpts by comparing detection of pitch changes that move away from the musically expected pitch (*musically unexpected*) to those that move toward or conform to the musically expected pitch (*musically expected*). If listeners are activating Western musical representations when they hear an excerpt as song, musically unexpected pitch changes should be more easily detected because they are incongruent with the expected melodic pattern, whereas musically expected pitch changes should fit well within the expected melody and therefore should be harder to notice. By contrast, if the excerpt does not activate Western musical representations because it is heard as speech, both pitch changes should be detected equally well because both pitch change types do not conform to the spoken contour. Likewise, if pitch change detection depends entirely on low-level acoustic characteristics, then discrimination performance should be comparable regardless of percept. By comparing perception of the same stimulus when it is heard as speech versus when heard as song, any differences in discrimination performance cannot be explained by low-level acoustic features, but rather by recruitment of high-level, domain-specific knowledge.

2. Experiment

2.1 Material and Methods

2.2 Participants

Forty-eight participants (24 musicians and 24 non-musicians) were recruited through the undergraduate psychology participant pool or through flyers and word-of-mouth communication. Participants completed demographic questionnaires after the experiment. Musicians (12 females) were 23.56 years of age on average (range: 18 – 56 years) and reported an average of 10.56 years of formal music training. Nineteen musicians were native English speakers (or learned another language simultaneously); others learned English as a second language before age 6 (Spanish =3,

Armenian =1, Russian & Hebrew = 1, Korean =1). Non-musicians (14 females) were 19.26 years old on average (range: 18 – 22) and reported 0.77 years of formal music training on average. Twenty-two non-musicians were native English speakers and others learned English before age 6 (n=1) or were Spanish-English bilinguals from birth (n=1). Ten additional participants were excluded from the study because of technical difficulties (n=1), ear infection or severe cold (n=2), or failure to follow instructions (e.g., responding “same” on every trial, not responding on more than 50% of trials, not attending to or misunderstanding the task; n=7). Previous studies from our lab indicated that 24 non-musicians were sufficient to observe that a single speech-to-song illusion reliably transformed to song at the group level (Vanden Bosch der Nederlanden et al., 2015). For this reason we chose to collect data from 24 participants in each group, which would provide us with a sufficient number of non-musicians perceiving transformations to song. We continued to run participants until 24 healthy musicians and non-musicians successfully completed the task.

2.3 Apparatus

Stimuli were presented through headphones at about 60 dB SPL. Participants’ computer keyboard responses were recorded using a custom script written with Psyscope X, Build 57 (Cohen, MacWhinney, Flatt, Provost, 1993) on an Apple Mac mini Intel core duo with OS X.

2.4 Stimuli

Twenty-four excerpts of natural speech were selected for this experiment. All of the excerpts were previously demonstrated to transform from speech to song after repetition, and were used with permission from the authors (Tierney et al., 2012). These stimuli were taken from online, open-access audiobook recording websites and consisted of speech segments from three male speakers. Relative to a phonetically similar corpus of non-transforming speech excerpts,

these transforming excerpts were reported to have significantly greater fundamental frequency stability and marginally greater rhythmicity as measured by the regularity of the intervals between stressed syllables (Tierney et al., 2012).

For each of the 24 excerpts, the location of the change (i.e., which syllable was altered) varied between stimuli. Eligible syllables had a more stable fundamental frequency, as it was difficult to detect a pitch change of one semitone when the syllable contained a pitch glide. We made pitch changes using Praat and exported them using the PSOLA overlap-add method in order to preserve formant frequency characteristics (Praat Software; Boersma and Weenink, 2001). The discrimination task was presented both at the beginning and at the end of each trial to capture performance when the excerpt was likely to be heard as speech (at the beginning) and as song (at the end).

To create pitch changes that either did or did not conform to Western musical scale expectations, we first estimated the perceived *sung contour*, or the melodic structure that listeners would presumably perceive when hearing the excerpt as song. In the same manner as Deutsch et al., 2011, the sung contour for each excerpt was ascertained by having a listener (C.V.B.D.N.) transcribe a vocal and musical instrument reproduction of each utterance after having heard it transform to song. Importantly, this rendition of the excerpt conformed to Western scale intervals. To verify each sung contour, seven participants gave informed consent to participate in a pilot study, in which they heard 10 repetitions of each original speech excerpt and then rated the similarity of a subsequent piano realization of the sung contour, using a rating scale from 1 = not similar to 5 = very similar. Piano renditions were rated as highly similar (average rating of 3 or above) or had at least two “very similar” ratings.

The average pitch for each syllable in the original excerpt was calculated using Praat's autocorrelation method over the voiced duration of a syllable. All pitch changes consisted of one-semitone shifts up or down in pitch from the average pitch of the original recorded syllable. Musically expected pitch changes moved the average pitch toward the sung contour, whereas musically unexpected pitch changes moved the pitch away from the sung contour by the same increment, as depicted in Figure 1. For example, if the average pitch of a syllable in the original excerpt was 195 Hz, approximately G3 on the musical scale, and the sung contour for that syllable was a G#3 (208 Hz), the musically expected pitch change raised the syllable one semitone to 209 Hz, while the musically unexpected change lowered it by one semitone to 186 Hz.

Figure 1: This figure illustrates an example of a musically expected and a musically unexpected pitch manipulation for the syllable “When” in the sentence fragment “When she comes she may

find.” Open diamonds represent the average pitch of each syllable of the spoken sentence, whereas asterisks illustrate the average pitch of each syllable as if it were sung according to the Western musical scale (called the sung contour). The grey circle represents the altered pitch of the syllable in the musically expected condition, when the direction of change conforms to the sung contour (asterisk). The open circle represents the altered pitch of the same syllable in the musically unexpected condition, when the direction of change moves away from the sung contour (asterisk). Note that the size is equivalent for musically expected and musically unexpected changes. Rhythmic information is not depicted here.

Procedure

After providing informed consent, participants were given verbal and written instructions by the experimenter. Instructions highlighted the order of the rating and discrimination tasks within a single trial. As in Deutsch et al., 2011, each trial began by presenting the full speech context (average length = 8604 ms, range 5483 – 15612 ms), which included the target excerpt. After the full context and a 2300 ms pause, we presented participants with one repetition of the speech segment (average length = 1333 ms, range 841 – 1799 ms) and asked them to provide a subjective rating of how song- or speech-like the excerpt sounded from 1 (“exactly like speech”) to 5 (“exactly like singing”) via button press. The next two repetitions comprised the discrimination task, in which participants indicated whether the pair of excerpts were the same (“S”) or different (“K”). Participants then continued to provide subjective ratings for each of the five subsequent repetitions (occurring with 2300 ms ISI). After the 8th repetition, pairs of excerpts were again presented as the final discrimination task. Thus for each trial, a participant heard 10 repetitions of the stimulus with a total of six iterations of the rating task and two iterations (containing four repetitions total) of the discrimination task (see Figure 2 for a depiction of a single trial).

Figure 2: The order of rating and discrimination tasks within a trial. Each trial began with the full context and was followed by 10 repetitions of the same speech segment taken from the full context of the speech sample. Importantly, a discrimination task was performed near the beginning to capture a spoken mode of listening and at the end of the repetition sequence for a sung mode of listening.

The discrimination task presented “same” trials, when the standard and comparison stimuli were identical ($N = 16$), “different” trials when the comparison contained a musically expected pitch change ($N = 16$), or “different” trials with a musically unexpected pitch change ($N = 16$). We used a greater proportion of “different” trials in order to have a sufficient number of musically expected and musically unexpected trials for calculating d' , which takes into account any response bias that may have resulted from the greater number of different trials. For a given trial, the initial and final discrimination pairs were independent, so the status of the initial discrimination pair (as containing a musically expected change, for example) did not predict the final repetition pair (it could be any of the three types of pairings). Participants provided responses for 24 trials administered within a single test block. Three quasi-random trial orders were assigned randomly to subjects using a Latin-square design, allowing each sentence to be presented with all three types of discrimination task pitch manipulations in a between-subjects fashion.

3. Results

For each participant and each pitch change type (musically expected or musically unexpected), discrimination (d') scores were calculated from the proportion of hits (“different” responses on change trials) and false alarms (“different” responses on same trials) assuming an independent-observation strategy, which is typical for a same-different task involving a standard and comparison stimulus (Macmillan & Creelman, 2005).¹

Not all excerpts were perceived as transforming from speech to song for all participants. Because our goal was to examine sensitivity to pitch manipulations when listeners heard an excerpt as both speech and as song, we only included in the analysis trials in which participants perceived a transformation from speech to song, as indicated by their subjective ratings (when they gave a rating of 1 or 2 to the initial rating task repetition *and* a rating of 3, 4, or 5 to the final rating task repetition; see Figure 1). For comparison, we also separately analyzed non-transforming trials in which listeners consistently heard the excerpt as not transforming from speech to song (i.e., either speech–speech or song–song). On average, about half of the trials for each participant were perceived as transforming (44% transforming, 55% stable), and a chi-square analysis of the number of times each individual excerpt was categorized as stable or transforming was non-significant, $\chi^2(23, N=997) = 17.096, p = .805$. Thus, none of the excerpts were more likely to be heard as either stable or transforming by all participants. Seven non-musicians and four musicians were excluded from this analysis because fewer than two trials remained in each cell of the design after list-wise deletion, a standard method for handling

¹ Because some participants had zero false alarms or a 1.0 hit rate, we applied a log-linear correction to avoid infinite values for d' by adding 0.5 to the numerator and 1 to the denominator (i.e., number of trials; Hautus, 1995).

missing data in an ANOVA. There were no main effects or interactions found for participants with other language experience, so language background was not included as a factor.

Sensitivity scores (d') for transforming trials were entered into a 2 x 2 x 2 (Group [musicians, non-musicians] x Position [initial, final] x Change Type [musically expected, musically unexpected]) mixed-design ANOVA. There was no main effect for group, $F(1,35) = 2.481, p = .124, \eta_p^2 = .066$, although musicians tended to have increased sensitivity in all conditions, and there were no interactions with group (p 's $> .3$), suggesting that musicians and non-musicians exhibited the same pattern of performance. No main effects were observed for position, $F(1, 35) = 2.804, p = .103, \eta_p^2 = .074$, or change type, $F(1, 35) = .892, p = .351, \eta_p^2 = .025$, but there was a significant interaction between position and change type, $F(1, 35) = 6.306, p < .05, \eta_p^2 = .153$. Two simple effects analyses performed separately for initial and final discrimination tasks revealed that, while sensitivity to pitch change types did not differ for the initial discrimination task, $F(1,35) = .745, p = .394, \eta_p^2 = .021$, sensitivity for musically unexpected changes was greater than musically expected changes during the final discrimination task, $F(1,35) = 4.372, p < .05, \eta_p^2 = .111$ (see Figure 3). In other words, musically expected pitch changes were harder to detect than musically unexpected pitch changes when excerpts were perceived as song, whereas no difference in sensitivity was observed for pitch changes during the initial discrimination task, when sentences were perceived as speech.

Figure 3: Participants' d' illustrates that only when the stimulus transforms to song do participants exhibit a difference in sensitivity for musically expected and musically unexpected trials. Stable trials show an enhancement in sensitivity overall, perhaps due to better encoding of exact pitch information or because of the musical representation of the stimulus. Error bars are within-subject standard error (Cousineau, 2005).

For comparison, we also conducted a separate analysis for *stable* trials, in which listeners perceived no transformation from speech to song. This analysis required list-wise deletion of four participants (2 musicians and 2 non-musicians) who had fewer than two trials in a given cell of the factorial design. Participants' d' scores were submitted to a 2 x 2 x 2 (Group [musicians, non-musicians] x Position [initial, final] x Change Type [musically expected, musically unexpected]) mixed design ANOVA. There was a significant main effect of group, $F(1, 42) = 4.911$, $p = .032$, $\eta_p^2 = .105$, with higher sensitivity overall for musicians than non-musicians, but there were no other interactions with group (p 's > .45). There was marginally better discrimination at the final than initial position of the trial, $F(1, 42) = 3.823$, $p = .057$, $\eta_p^2 = .083$, (Figure 3). We found no main effect for change type, $F(1, 42) = .475$, $p = .495$, $\eta_p^2 = .011$, and no interaction of change type and position, $F(1, 42) = .149$, $p = .702$, $\eta_p^2 = .004$, indicating that the marginal increase in sensitivity from initial to final position was the same for musically

expected *and* musically unexpected pitch changes. Thus, when participants do not report the illusory transformation from speech to song, no difference in the pattern of sensitivity is found for musically expected and musically unexpected pitch changes, only a general increase in sensitivity overall.

4. Discussion

The present results provide support for the notion that listeners can recruit music-specific cognitive representations of pitch structure while listening to speech. When listeners indicate that they hear a recorded utterance as speech, they are equally sensitive to pitch changes that do or do not conform to Western musical scales. By contrast, when listeners indicate that they hear a recorded speech utterance as song, they are sensitive to pitch changes that violate musical expectations, but they miss changes that conform to their expectations. If, as a result of repetition, domain-general factors such as enhanced encoding or the reallocation of attention toward pitch were the only factors driving discrimination, then sensitivity would be greater for final compared to initial discrimination tasks regardless of percept (i.e., speech vs. song) or manipulation type (i.e., musically expected vs. musically unexpected). Enhanced encoding due to repetition may explain why sensitivity increases with repetition for excerpts that do not transform to song (stable trials), but domain-general factors alone cannot explain why sensitivity is reduced for pitch changes that conform to the musically expected contour during transforming trials. Only when repetition leads to a change in percept from speech to song do we see evidence that listeners are more sensitive to musically unexpected than expected pitch changes.

This study is the first to use an objective perceptual task to directly demonstrate that both musicians and non-musicians recruit music-specific knowledge when listening to an ambiguous stimulus that can be heard as speech or song, and it is consistent with previous research using a

vocal production task from Deutsch et al. (2011). Together these studies demonstrate that a single naturally occurring utterance can activate either a speech or music mode of listening, and that this listening mode influences pitch processing above and beyond low-level acoustic features of the stimulus.

Musicians and non-musicians both exhibited the facilitative and interfering effects of musical key knowledge, but musicians did display an overall trend toward better discrimination than non-musicians. Musicians were marginally more sensitive to pitch changes during transforming trials and significantly more sensitive than non-musicians for stable trials. These results are consistent with a wealth of previous research that finds enhanced pitch processing for musicians compared to non-musicians (e.g., Schön, Magne & Besson, 2004). Yet, despite average differences in sensitivity, the same pattern of results was found for musically expected and unexpected pitch manipulations, indicating that the music-specific processes elicited in this task do not depend on formal music training but can arise from implicit enculturation to Western music (e.g., Lynch, Eilers, Oller, & Urbano, 1990).

Implicit musical knowledge differentially affected listeners' sensitivity to musically expected and unexpected pitch changes when excerpts were perceived as transforming to song. When a pitch change moved toward a musically expected pitch, it presumably fit well with the musical expectations of Western listeners, making it harder to detect, even though the same magnitude of the physical change was readily detected for musically unexpected pitch changes. Interestingly, sensitivity to pitch changes generally increased on transforming as well as stable trials, suggesting that, in general, repetition enhances pitch encoding. We anticipated that activation of Western musical representations would enhance sensitivity to musically unexpected pitch changes, but instead we observe that activation of a musical percept paradoxically

interferes with or inhibits the low-level benefit of repetition for musically expected pitch changes. This result is in fact consistent with prior studies, which report that adults are much worse at detecting pitch changes that conform to musical expectations (e.g., within-key or within-harmony) than changes that violate expectations (out-of-key changes; Trainor & Trehub, 1992). This pattern is also consistent with the differential use of pitch information in speech and song (Zatorre & Baum, 2012). Top-down knowledge may inhibit low-level frequency change detection, or, alternatively, low-level details may not be available to the listener once the listener perceives a melodic percept as predicted by Reverse Hierarchy Theory (Hochstein & Ahissar, 2002). According to this theory, bottom-up processing of low-level stimulus details occurs first, followed by conscious perception of high-level categorical information (e.g., a melody's structure or meaning). Once listeners perceive a whole object in this manner, they can traverse down the perceptual hierarchy to selectively attend to and access an object's constituent parts (e.g., a particular note's frequency).

Despite identical acoustic characteristics, higher-level knowledge of the perceived domain influenced pitch sensitivity for a single stimulus, arguing against a purely domain-general, low-level account (Zatorre & Gandour, 2008). Of course, this is not to say that domain-general factors are not involved in an illusory transformation to song. While the physical characteristics did not change with repetition, stimuli may be perceived as transforming to song because repetition changes the salience of specific acoustic features (e.g., pitch stability, Falk et al., 2014; Margulis, 2013). A combination of domain-specific and domain-general processes likely contribute to this illusion, with music-specific knowledge being recruited or a musical percept activated depending on the weighting of low-level acoustic elements within the excerpt as a result of repetition. Sensitivity to different pitch manipulations, however, was context-

dependent, similar to auditory object identification studies in which objects that are incongruent with the auditory context are more readily detected than those that are congruent (Krishnan, Leech, Aydelott, & Dick, 2013). Further studies should examine the recruitment of music- and language-specific knowledge at different developmental stages, determining how and when children differentiate between music and language based on contextual demands or the acoustic characteristics unique to each domain.

Using a novel behavioral paradigm, we have provided evidence that knowledge for speech and song can be recruited separately even when low-level characteristics are held constant, suggesting that pitch representations are at least partially domain-specific and are not completely tied to low-level stimulus features. This finding paves the way for future studies to further tease apart high- and low-level contributions by examining neural and behavioral correlates of music and language processing.

Author Contributions: All authors contributed to the design of the study, C.V.B.D.N. performed all of the data collection and carried out analyses under the supervision of E.E.H. and J.S.S. Drafts of the manuscript were developed by C.V.B.D.N and critical revisions were provided by E.E.H. and J.S.S. All authors approved the final version of the manuscript.

Acknowledgments: The authors would like to thank Adam Tierney for the use of the auditory illusions from speech to song and to thank LibriVox and audiobooksforfree.com, from whom the speech excerpts were originally obtained. Data collection was completed with support from a grant from the National Science Foundation grant number BCS-1052718 awarded to E.E.H.

References

- Boersma, P. & Weenink, D. (2001). Praat, a system for doing phonetic by computer. *Glott International* 5(9/10), 341-345.
- Chomsky, N. (1959). A review of B. F. Skinner's verbal behavior. *Language*, 35(1), 26–58.
- Cohen J.D., MacWhinney B., Flatt M., and Provost J. (1993). PsyScope: A new graphic interactive environment for designing psychology experiments. *Behavioral Research Methods, Instruments, and Computers*, 25(2), 257-271.
- Cosmides, L., & Tooby, J. (1994). Origins of domain specificity: The evolution of functional organization. *Mapping the mind: domain specificity in cognition and culture* (pp. 85–116). Cambridge, UK: Cambridge University Press.
- Cousineau, D. (2005). Confidence intervals in within-subjects designs: A simpler solution to Loftus and Masson's method. *Tutorials in Quantitative Methods for Psychology*, 1(1), 42-45.
- 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). *Phantom Words, and Other Curiosities*. (Philomel Records, La Jolla) Compact Disc and Booklet; Track 22.
- Deutsch, D., Henthorn, T., & Lapidis, R. (2011). The illusory transformation of speech to song. *Journal of the Acoustical Society of America*, 129 (4), 2245-52.

- Falk, S., Rathcke, T., & Dalla Bella, S. (2014). When speech sounds like music. *Journal of Experimental Psychology: Human Perception and Performance*. Advance online publication. <http://dx.doi.org/10.1037/10036858>
- Fedorenko, E., Patel, A.D., Casasanto, D., Winawer, J., & Gibson, E. (2009). Structural integration in language and music: Evidence for a shared system. *Memory & Cognition*, 37(1), 1-9.
- Fodor, J. A. (1983). *The modularity of mind: An essay on faculty psychology*. MIT Press.
- Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science*, 298, 1569 – 1579.
- Hautus, M.J. (1995). Corrections for extreme proportions and their biasing effects on estimated values of d' . *Behavior Research Methods, Instruments, & Computers*, 27(1), 46-51.
- Hochstein, S. & Ahissar, M. (2002). View from the top: Hierarchies and reverse hierarchies in the visual system. *Neuron*, 36, 791-804.
- Joanisse, M. F. & Gati, J. S. (2003). Overlapping neural regions for processing rapid temporal cues in speech and nonspeech. *NeuroImage*, 19, 64-79.
- Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: evidence for a domain general learning mechanism. *Cognition*, 83, B35-B42.
- Krishnan, S., Leech, R., Aydelott, J., & Dick, F. (2013). School-age children's environmental object identification in natural auditory scenes: Effects of masking and contextual congruence. *Hearing Research*, 300, 46-55.
- Levitin, D. J., & Menon, V. (2003). Musical structure is processed in “language” areas of the brain: a possible role for Brodmann Area 47 in temporal coherence. *NeuroImage*, 20(4), 2142–2152.

- Lynch, M. P., Eilers, R. E., Oller, K. D., & Urbano, R. C. (1990). Innateness, experience, and music perception. *Psychological Science*, 1(4), 272-276.
- MacMillan, N.A. & Creelman, C.D. (2005). Detection Theory: A User's Guide – 2nd Edition. Lawrence Erlbaum Associates.
- Margulis, E. H. (2013). Repetition and emotive communication in music versus speech. *Frontiers in Psychology*, 4, 1-4.
- 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.
- Merrett, D. L., Peretz, I., & Wilson, S. J. (2014). Neurobiological, cognitive, and emotional mechanisms in Melodic Intonation Therapy. *Frontiers in Human Neuroscience*, 8, 1-11.
- Newport, E. L. (2011). The modularity issue in language acquisition: A rapprochement? Comments on Gallistel and Chomsky. *Language Learning and Development*, 7, 279-286.
- Patel, A. D., Gibson, E., Ratner, J., Besson, M. & Holcombe, P. J. (1998). Processing syntactic relations in language and music: And event-related potential study. *Journal of Cognitive Neuroscience*, 10(6), 717-733.
- Patel, A. D. (2011). Why would musical training benefit the neural encoding of speech? The OPERA hypothesis. *Frontiers in Psychology*, 2, 1-14.
- Peretz, I., Ayotte, J., Zatorre, R. J., Mehler, J., Ahad, P., Penhune, V. B., & Jutras, B. (2002). Congenital Amusia: A disorder of fine-grained pitch discrimination. *Neuron*, 33, 185-191.
- Peretz, I. & Hyde, K. L. (2003). What is specific to music processing? Insights from congenital amusia. *Trends in Cognitive Sciences*, 7(8), 362 – 367.

- 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. *The Journal of Neuroscience*, 31(10), 3843–3852.
- Saffran, J.R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, 274, 1926-1928.
- Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, 70(1), 27–52.
- Schön, D., Magne, C., Besson, M. (2004). The music of speech: Music training facilitates pitch processing in both music and language. *Psychophysiology*, 41, 341-349.
- Tervaniemi, M. & Hugdahl, K. (2003). Lateralization of auditory-cortex functions. *Brain Research Reviews*, 43, 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. (2012). Speech versus song: Multiple pitch-sensitive areas revealed by a naturally occurring musical illusion. *Cerebral Cortex*, 23, 249-254.

- Trainor, L. J., & Trehub, S. E. (1992). A comparison of infants' and adults' sensitivity to Western Musical Structure. *Journal of Experimental Psychology: Human Perception and Performance*, 18(2), 394–402.
- 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. (2015). Everyday musical experience is sufficient to perceive the speech-to-song illusion. *Journal of Experimental Psychology: General*, epublication ahead of print.
- Zatorre, R., Evans, A., Meyer, E., & Gjedde, A. (1992). Lateralization of phonetic and pitch discrimination in speech processing. *Science*, 256(5058), 846–849.
- 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. J., & Gandour, J. T. (2008). Neural specializations for speech and pitch: moving beyond the dichotomies. *Philosophical Transactions of the Royal Society: Biological Sciences*, 363(1493), 1087–1104.