Evidence for two mechanisms to account for the speech to song illusion, the verbal transformation effect, and the sound to music illusion.

Michael S. Vitevitch
ORCID: 0000-0002-1209-0838
University of Kansas

Elizabeth R. Phillips
University of Kansas

Evan A. Norkey
ORCID: 0000-0003-4682-5211
University of Kansas

Anisha Kodwani
ORCID: 0000-0001-5173-2787
University of Kansas

Correspondence should be addressed to:

Michael S. Vitevitch, Ph.D.
Spoken Language Laboratory
Department of Psychology
1415 Jayhawk Blvd.
University of Kansas
Lawrence, KS 66045
e-mail: mvitevit@ku.edu
ph: 785-864-9312

Abstract

Introduction: Five studies examined the speech to song illusion, the verbal transformation effect, and the sound to music illusion in order to determine if they were distinct phenomena and to assess if they could be accounted for by a single perceptual/cognitive mechanism.

Methods: In Study 1, word lists varying in length from 1 word (as often used to study the verbal transformation effect) to 4 words (as often used to study the speech to song illusion) were presented to participants for 4 minutes to investigate the percepts that were elicited. In Study 2 participants were asked to indicate YES/NO if they experienced the speech to song illusion when listening to word-lists modified by a vocoder. In Studies 3-5 participants were asked to click a button as soon as the shift in percept occurred from speech (or sound) to a music-like percept to assess the time-course of the speech to song (or sound to music) illusion.

Results: Study 1 shows that the verbal transformation effect and the speech to song illusion elicit similar percepts. In Study 2 participants indicated that the speech-like stimuli elicited the speech to song illusion more than the noise-like stimuli. In Studies 3-5 similar time-courses were observed for the speech to song illusion and the sound to music illusion.

Discussion: Previous, single-mechanism accounts of the speech to song illusion are discussed, but none of them adequately account for all of the results presented here. A new model is proposed that appeals to both a perceptual/"lower-level" mechanism and a cognitive/"higher-level" mechanism.

Keywords: speech to song illusion; node structure theory; verbal transformation effect; sound to music illusion

Introduction

In the auditory domain there are several illusions that occur when a stimulus has been presented repeatedly, namely, the verbal transformation effect, the speech to song illusion, and the sound to music illusion. In the verbal transformation effect, a single word, like *flame*, is presented repeatedly, resulting in listeners initially reporting that they hear the word *flame*, but after a number of repetitions, the percept changes and they report hearing the word *blame*, or *lame,* or *fame* (Warren & Gregory, 1958; see Kaminska & Mayer, 2002 for transformations of non-speech sounds). In the speech to song illusion, a phrase or list of words is presented repeatedly. Initially, listeners report that the phrase sounds as if it is being spoken, but after several repetitions, the percept changes and listeners report that the phrase sounds as if it is being sung (Deutsch, Henthorn & Lapidis, 2011). In the sound to music illusion, nonspeech, environmental sounds, such as water dripping or a shovel being dragged across a rock, are initially reported as sounding like those environmental sounds. After several repetitions, the percept changes and listeners report that the environmental sounds have taken on a music-like quality (Simchy-Gross & Margulis, 2018).

Although these auditory illusions are all evoked by repetition of the stimulus (whether it be a word, phrase, or non-speech sound), they are often examined separately, as if they were independent or distinct phenomena. We sought in the present set of studies to examine whether these auditory illusions are actually distinct phenomena or simply appear to be different due to the different stimuli employed (speech vs. non-speech sounds), the variation in the tasks that are typically used to examine each of them, or the specific responses that participants are typically asked to report in investigations of each illusion. In exploring the distinctiveness of these three illusions, we also considered if a common perceptual or cognitive mechanism was responsible

for these auditory illusions, or if more than one perceptual or cognitive mechanism might play a role in evoking these auditory illusions.

One account of these three illusions proposes that repetition is the sole mechanism that causes each illusion (Margulis, 2013; Margulis & Simchy-Gross, 2016; Rowland et al., 2019). Clearly, repetition of the stimulus plays a role in the speech to song illusion, as well as the verbal transformation effect, and the sound to music illusion, but repetition alone is not a sufficient explanation for how or why any of these auditory illusions occur, or for why one illusion occurs instead of another. For example, a simple repetition account does not explain why repetition of a list of words results in elicitation of the speech to song illusion rather than the verbal transformation effect, which occurs with the repetition of single words. Mechanisms beyond repetition alone may be required to adequately account for the differences among these illusions.

In addition to not adequately explaining why one illusion occurs instead of another, the repetition account—as initially proposed by Deutsch et al. (2011) to account for the speech to song illusion—contradicts what we know about how language is processed. Deutsch et al. (2011; p. 2251) hypothesized that:

> …in listening to the normal flow of speech, the neural circuitry underlying pitch salience is somewhat inhibited, perhaps to enable the listener to focus more on other characteristics of the speech stream that are essential to meaning, i.e., consonants and vowels. We can also hypothesize that exact repetition of the phrase causes this circuitry to become disinhibited, with the result that the salience of the perceived pitches is enhanced.

Psycholinguistic evidence suggests that pitch is an essential acoustic feature used to understand the various languages of the world. In tone languages, like Mandarin, variation in pitch is used to distinguish the meaning of syllables that contain the same phonemes. The classic example in Mandarin is of the syllable /ma/. With a high, level pitch (tone 1), /ma/ means *mother*. With a low pitch that rises to a higher pitch (tone 2), /ma/ means *hemp*. When the pitch

starts high, dips low, and rises again (tone 3), /ma/ means *horse*. When the pitch starts high and drops sharply (tone 4), /ma/ means *scold*. Given the important role that pitch plays in tone languages like Mandarin (and in pitch-accent languages like Japanese and stress-timed languages like English), it is unclear why the neural circuitry used to process pitch would be inhibited as suggested by Deutsch et al. (2011). Indeed, it is important to note that the speech to song illusion has been observed in Mandarin (Zhang, 2011; Jaisin et al. 2016) as well as in English, further undermining the repetition account.

Although positing that a single mechanism—such as repetition (e.g., Rowland, Kasdan & Poeppel, 2019)—underlies all of these auditory illusions makes for a parsimonious account, there are two facts that raise the possibility that there may be more than one mechanism responsible for these illusions. First, accounts of other perceptual and cognitive processes posit that multiple mechanisms are involved in processing. For example, in color perception, there are two well-studied mechanisms—the trichromatic theory (Young, 1802) and the opponent-process theory (Hering, 1872)—that are both required to fully explain various phenomena related to color perception, including color blindness and color after-images. As with color perception, more than one mechanism may be required to fully account for *language-based auditory illusions* (such as the verbal transformation effect and the speech to song illusion) and for *music-related auditory illusions* (such as the speech to song illusion and the sound to music illusion)[1].

Second, music and language processing may have emerged evolutionarily at different times and thus may have multiple mechanisms underlying them. Consider that songbirds and whales communicate with song-like vocalizations, but that scant evidence exists in animal

[1] These categories are intended merely as a narrative aid in the reporting of our investigations, not as a scientifically established taxonomy. Indeed, the verbal transformation effect and the speech to song illusion can be elicited with made-up words that do not exist in a given language (e.g., Castro et al., 2018; Shoaf & Pitt, 2002).

communication systems for anything that resembles the systems of phonology, morphology, semantics, syntax, etc. found in human languages (Hauser et al., 2014; see also Haiduk & Fitch, 2022; Eleuteri, V., et al., 2022). There is also evidence for distinct cortical pathways for processing music and speech (Norman-Haignere, Kanwisher & McDermott, 2015), further suggesting that more than one mechanism may underlie these three auditory illusions that emerge when the stimulus is repeated.

In the present set of studies (see Table 1), we examined the distinctiveness of these three auditory illusions using a variety of methodologies. In Study 1 we presented listeners with lists that contained 1-4 words, and asked them to report any changes they experienced in the percept to examine the distinctiveness of the verbal transformation effect (typically evoked by a single word) and the speech to song illusion (typically evoked by multi-word phrases). The open-ended reporting method allowed participants to report percepts that might typically only be reported for one or the other auditory illusion due to the constraints imposed by the task typically used to examine the verbal transformation effect, or by the task typically used to examine the speech to song illusion.

Table 1. Summary of the methods employed in the five studies in this report.

Study	Illusions examined	Stimulus	Task
1	Verbal Transformation Effect Speech to Song Illusion	Lists containing 1-4 words	Report any changes in percept
2	Speech to Song Illusion Sound to Music Illusion	Vocoded speech	Did you experience the speech to song illusion?
3	Speech to Song Illusion	Word-lists varying in neighborhood density	Click when percept changes from speech to song.
4	Sound to Music Illusion	4 environmental sounds	Click when percept changes from sound to music.
5	Sound to Music Illusion	4 “emotional” sounds	Click when percept changes from sound to music.

In Study 2 we examined the distinctiveness of the speech to song illusion and the sound to music illusion by presenting listeners with a vocoded stimulus that could be perceived as either speech or as noise, and simply asked participants if they experienced the speech to song illusion (i.e., yes/no). If the speech to song illusion and the sound to music illusion are distinct phenomena, then only speech-like stimuli should elicit a change in the percept from speech to song.

In Study 3 we attempted to determine if the mechanism that underlies the speech to song illusion is cognitive/"higher-level" in nature or perceptual/"lower-level" in nature by manipulating the characteristics of the words that were repeated, and asking participants to press a button as soon as the percept changed from speech to song. By measuring the time-to-transform from speech to song, we were also able to assess the time-course of the speech to song illusion in addition to localizing the mechanism (i.e., higher- or lower-level) that underlies the illusion.

In Study 4 we examined further the distinctiveness of the speech to song illusion and the sound to music illusion. In this study we used 4 environmental sounds and the same time-to-transform paradigm employed in Study 3 to determine if the sound to music illusion had a different time-course than the speech to song illusion.

Finally, in Study 5 we again examined the distinctiveness of the speech to song illusion and the sound to music illusion by considering how emotional stimuli might influence the two illusions. A previous study of the speech to song illusion found that words varying in emotional arousal did not differentially influence song-likeness ratings in the speech to song illusion (Vitevitch, Ng, Hatley & Castro, 2021). In the present study, we used "emotional sounds" to see

if emotion had an influence on the sound to music illusion that differed from the (lack of) effect of emotion in the speech to song illusion.

Together, the results of these five studies will enable us to determine if the three auditory illusions are as distinct as previously assumed. Further, the results of these five studies will provide important insight into the mechanism (or mechanisms) that might underlie these three auditory illusions.

Study 1

In this study we focused on the two language-based illusions, namely the verbal transformation effect and the speech to song illusion. We categorized the verbal transformation effect as language-based because a *spoken word* is typically repeated, and its percept "transforms" into a different spoken word. We categorized the *speech* to song illusion as language-based because a phrase or list of words is repeated, and the percept changes from the phrase/list of words being spoken to being sung.

Instead of using the repetition account of the speech to song illusion to make predictions, we used an alternative account of the illusion that appeals to the language processing model known as Node Structure Theory (*NST*; MacKay, 1987). Not only does NST account for the speech to song illusion (Castro, Mendoza, Tampke & Vitevitch, 2018; Mullin, Norkey, Kodwani, Vitevitch & Castro, 2021; Vitevitch, Ng, Hatley & Castro, 2021), but it also accounts for the verbal transformation effect (MacKay, Wulf, Yin & Abrams, 1993; Shoaf & Pitt, 2002).

In Node Structure Theory (MacKay, 1987), nodes represent phonemes, syllables, words, and other types of linguistic information. Links connect nodes such that phoneme nodes connect to syllable nodes, syllable nodes connect to lexical nodes, etc. (see Figure 1). (Note that the type of network formed by the nodes and links in NST is very different from the type of complex

networks described in Vitevitch, 2022.) During speech perception, incoming acoustic-phonetic information *primes* (similar to spreading activation in other models) phonological nodes based on the extent to which the nodes match the input. When a node accumulates enough priming to surpass an activation threshold, the node is *activated*, bringing to conscious awareness the information represented by that node.

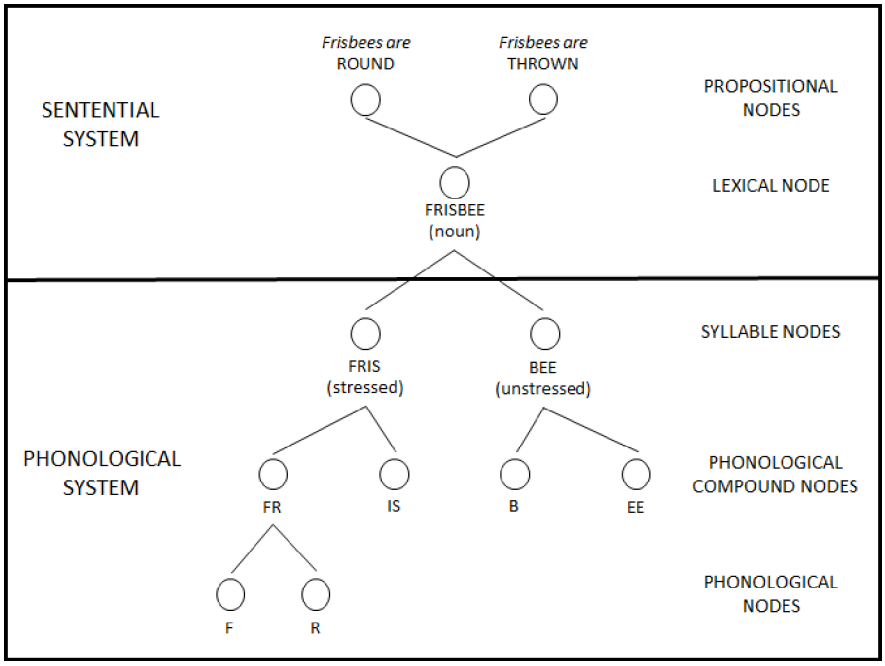

Figure 1. Nodes representing various types of linguistic information for the word *frisbee*. Additional higher-level and lower-level nodes described in Node Structure Theory have been omitted to simplify the image.

Presentation of a word or phrase initially primes and activates lexical nodes associated with those words, bringing to conscious awareness the information associated with the word (and a speech-like percept). With repeated activation of the same lexical nodes, *satiation* occurs, resulting in the lexical nodes being temporarily unable to accumulate priming and be activated and thus in the inability to retrieve information associated with that word/lexical node (Vitevitch et al., 2021).

In the case of the speech to song illusion, additional presentations of the stimulus continue to prime the syllable nodes. Because syllables continue to receive priming, the syllable nodes make salient the rhythmic pattern in the repeated phrase, resulting in a song-like percept. Note that syllables are widely recognized as a unit of rhythmic structure in speech (e.g., Cutler, 1991; Fujii & Wan, 2014; Jackendoff, 2009; Ramus, Nespor & Mehler, 1999). (Alternatively, one may consider the theory of phonology known as Beats and Binding (Dziubalska-Kołaczyk, 2002), which eschews syllables, but nevertheless emphasizes a rhythmic skeleton that contains regularly recuring beats (typically a vowel) and non-beats (always a consonant) that are "bound" together in words.) Several studies of the speech to song illusion have tested and confirmed many of the predictions derived from NST (Castro et al., 2018; Mullin, et al., 2021; Vitevitch et al., 2021).

In the case of the verbal transformation effect, the lexical node associated with the repeated word satiates. Satiation of the lexical node associated with the input gives another lexical node that is similar to the input the opportunity to be primed and ultimately activated by the repeated stimulus, bringing to conscious awareness another word (MacKay, et al., 1993). Several studies have tested predictions derived from NST to account for the verbal transformation effect (MacKay, Wulf, Yin & Abrams, 1993; Shoaf & Pitt, 2002).

In most studies of the verbal transformation effect 1 word is repeated (e.g., Shoaf & Pitt, 2002), but occasionally 2 words have been used as the stimulus (e.g., Kaminska & Mayer, 2002). In the speech to song illusion short sentences or phrases extracted from sentences are typically used as stimuli (e.g., Deutsch et al., 2011), but short lists of words have also been shown to elicit the illusion (Castro et al., 2018). In Experiments 5 and 6 of Castro et al. (2018) the ideal number of words needed to evoke the speech to song illusion was found to be 3-4 words. In the present

study we repeatedly presented listeners with lists that contained 1 to 4 words and asked them to report the percepts and any changes in percepts that they might experience.

In most studies of the verbal transformation effect participants are typically asked to simply state what word (or nonword) they are perceiving, whereas in most studies of the speech to song illusion, participants are typically asked to indicate if they experienced the song-like percept or to rate the strength of the illusion (see also Mullin, et al., 2021). With the important exception of Kaminska and Mayer (2002), participants are not typically given the opportunity to report any other percepts they may experience in a study of the verbal transformation effect (or in a study of the speech to song illusion). In the present study we used a more open-ended methodology like that used by Kaminska and Mayer (2002), and asked participants to report the percepts and any changes in percepts that they might experience, providing participants the opportunity to report on a wider range of percepts should they be experienced.

In previous studies of the speech to song illusion (e.g., Castro et al, 2018), the phrase is repeated 10 times. In the present study we instead repeated each list for 4 minutes (as in Kaminska & Mayer, 2002), which is closer to the presentation times used to elicit the verbal transformation effect. Thus, the present study combined methodologies often used to examine independently the verbal transformation effect and the speech to song illusion. Combining the methodologies and stimuli typically employed to examine independently the verbal transformation effect and the speech to song illusion allowed us to explore what percepts (if any) might lie "between" the two auditory illusions. The combination of methods and stimuli also allowed us to examine if there is any overlap in the percepts experienced in the two auditory illusions, which would suggest that the two illusions may not be as distinct as previous descriptions of these phenomena might lead one to believe.

Although the present study is exploratory in nature, we made a couple of tentative predictions. First, if the repetition account brings about both the verbal transformation effect and the speech to song illusion, then the number of illusory changes reported should remain relatively constant over time. That is, the number of illusory changes reported after 1 minute of repetition should be about the same number of illusory changes reported after 4 minutes of repetition. In contrast, if the NST account of the verbal transformation effect and the speech to song illusion brings about both illusions, then one would expect more illusory changes reported after 4 minutes of repetition than after 1 minute of repetition, because the satiation of nodes takes some time to occur. Also, many more nodes would be satiated later in the session than earlier in the session, perhaps leading to additional transformations or changes in percepts.

Second, if the verbal transformation effect and the speech to song illusion are indeed distinct illusions, then distinct percepts should be reported for each list length (i.e., 1 to 4 words in the list). That is, as the number of words in the list increases, listeners would report fewer percepts related to the verbal transformation effect (e.g., a different word, non-words), and instead report more percepts typically experienced in the speech to song illusion (e.g., pitch, rhythm). Alternatively, if the illusions are not completely independent, the same percepts may be reported for shorter word lists and for longer word lists.

Methods

All of the studies reported here were approved by the Institutional Review Board at the University of Kansas. Study 1 was initiated prior to March 2020. However, the restrictions imposed in the United States to restrict the spread of the COVID-19 virus after that date prevented us from collecting additional data in-person, resulting in a smaller than desired sample size.

Participants: Eighteen undergraduate students over the age of 18 years received participation credit for an Introductory Psychology class as compensation for their participation in this experiment. Participants were all native speakers of English, reported no speech or hearing disorders, and provided written informed consent. Note that data from 1 participant was not included in the analyses due to their failure to follow instructions during the experimental session.

Materials: The words used in the present study were the 28 words with dense phonological neighborhoods (Vitevitch & Luce, 2016) that were previously used in Castro et al. (2018), Soehlke, Kamat, Castro & Vitevitch (2022), and Vitevitch, Stamer & Sereno (2008), because these words have been shown to elicit the speech to song illusion. As originally reported in Vitevitch et al. (2008), the dense words had a mean of 11.71 (*sd* = 1.58) phonologically similar words. The bisyllabic words had stress on the first syllable (i.e., a strong-weak stress pattern), and were recorded by a female, native English speaker at a normal speaking rate. Recordings were made in an IAC sound-attenuated booth using a high-quality microphone onto a digital recorder at a sampling rate of 44.1 kHz. The words were edited into individual sound files using Sound Edit 16 (Macromedia, Inc.), then concatenated to create the lists used in the present study. None of the words in a list were phonological neighbors of another word in the list. The minimum pitch for the dense words = 161.60 Hz (*sd* = 53) and the maximum pitch for the dense words = 309.82 Hz (*sd* = 109).

Table 2. The words used as stimuli in Study 1.

	Set A	**Set B**	**Set C**
1 word	paddle	candle	valley
2 words	muscle body	dairy meter	leather babble
3 words	cattle berry mayor	bubble money ladder	banner candle worry
4 words	hurry puddle lighter shallow	furry mayor leather tackle	battle polar candy lever

Note: A given participant heard the four word-lists varying in length from Set A, or Set B, or Set C. The order of the word-lists varying in length was randomized for each participant.

Procedure: Each participant heard a list of words that contained 1 word, a list of words that contained 2 words, a list of words that contained 3 words, and a list of words that contained 4 words. Each list was repeated for 4 minutes regardless of the number of words in the list. The order of presentation for each list was randomized for each participant. Three different sets of words were used (labeled Set A, Set B, and Set C in Table 2) in order to generalize any effects that might be observed.

Participants were instructed to report immediately any illusory changes they perceived. Specifically, each participant was told:

> *In this lab we study auditory illusions. Auditory illusions are similar to optical illusions. In optical illusions, you see something that does not reflect physical reality. In auditory illusions, you hear something that does not reflect physical reality. Today you will be listening to a series of recordings containing repeated words. As you listen, please indicate out loud every time you perceive a change in the words you are hearing, as well as specifically what changed. You may hear changes in pitch, rhythm, or a word itself may change to something different. You may hear different types of changes other than those I*

just listed. You may hear no changes at all. Regardless, it is important that if you perceive a change, you should tell me right away. Please do not take off your headphones while you are telling me the change you heard. At the end of each list, I may ask you to clarify some of the statements you made, so please do not proceed to the next recording until I indicate that you should do so. After we finish discussing each recording, you will press the space bar to proceed to the next recording. Do you have any questions?

Participants wore a set of Beyerdynamic DT 100 headphones. Stimulus presentation was controlled by an iMac computer running PsyScope 1.2.2 (Cohen, MacWhinney, Flatt, Provost, 1993). The display on the monitor was blank with the exception of the number 1, 2, 3, or 4 appearing in the top right corner of the screen to provide to the research assistant an approximate time at which a report was made by a participant.

A research assistant recorded the change reported by the participant in the participant's own words and documented the minute within which the report occurred. At the end of a given word-list and before proceeding to the next word-list, the research assistant conferred with the participant to clarify any of their reports if needed.

Results

The illusory changes reported by the 17 participants who followed instructions were categorized using several of the categories originally developed by Kaminska and Mayer (2002): Pitch, Extra Sound, Separation of Elements, Rhythm, Volume, Emphasis/Stress, Rate, and Clarity. Several additional categories were created to accommodate reports that did not fit into one of the categories developed by Kaminska and Mayer (2002). We created a category specifically for transformations to a real word in English (word) and a category for transformations to a nonword (nonword). Note that we did not assess the phonotactic legality or probability of the nonwords (Vitevitch & Aljasser, 2021).

We also created a category for a new phenomenon we observed, which we have named *patternization*. In contrast to VTE, where a word changes from one word to another word (or nonword) to yet another word (or nonword), in patternization two versions of a given word or word list alternate with each other. For example, the word "candle" might alternate with the nonword "cando," becoming "candle, cando, candle, cando…".

Two trained research assistants independently categorized the responses. Over 90% agreement was obtained, and any discrepancies were resolved by consensus.

Figure 2 shows the total number of illusory changes reported during each minute for each word-list varying in length. A paired *t*-test was used to compare the number of illusory changes reported in minute 1 to the number of illusory changes reported in minute 4. A paired *t*-test was used because the 1-word list presented in minute 1 was also presented in minute 4; the 2-word list presented in minute 1 was also presented in minute 4; etc. The results show that more illusory changes were reported in minute 4 (*mean* = 30.0, *sd* = 12.03) than in minute 1 (*mean* = 16.25, *sd* = 9.74; $t(3) = 9.57$, $p = .0024$).

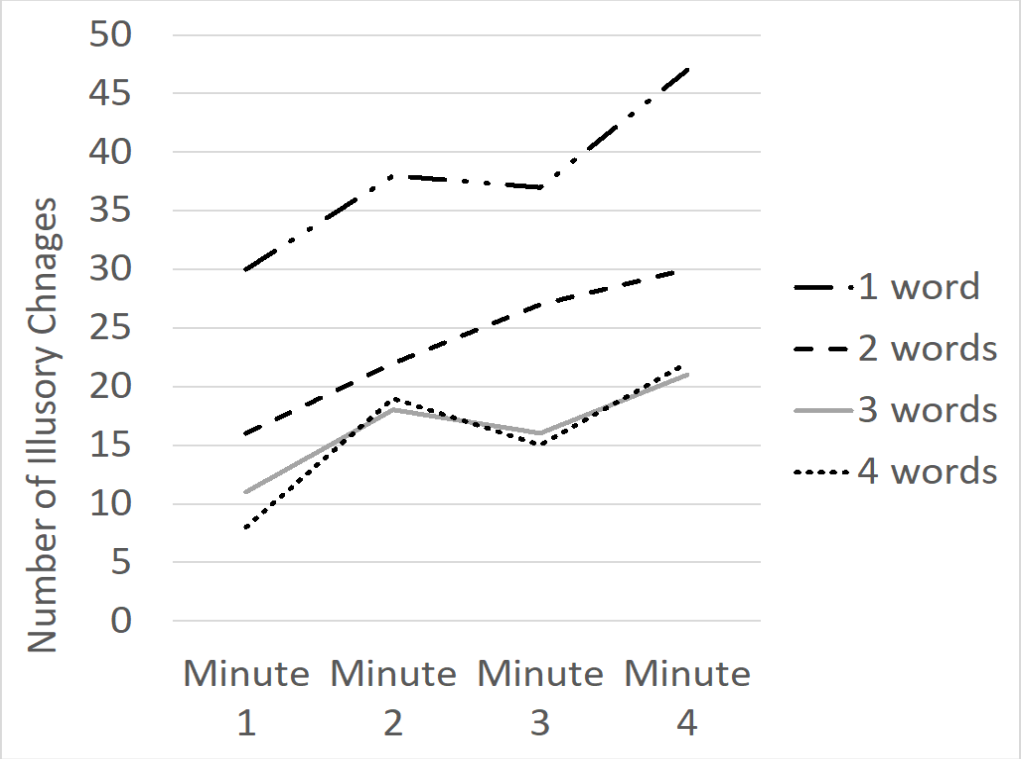

Figure 2. The number of illusory changes reported (from 17 participants) during each minute that a list was repeated.

Turning now to the types of illusory changes reported, Table 3 shows the number of illusory changes that are typically reported in studies of the verbal transformation effect (a different word or a non-word), the number of illusory changes that are typically reported in studies of the speech to song illusion (change in pitch or rhythm), and the number of patternizations for the number of words in the list.

Table 3. The total number of illusory changes typically associated with the verbal transformation effect (words, non-words), the number of illusory changes that are typically associated with the speech to song Illusion (pitch, rhythm), and the number of patternizations reported by 17 participants.

	Verbal Transformation Effect		**Speech to Song Illusion**		
	Words	**Non-words**	**Pitch**	**Rhythm**	**Patternization**
1 word	8	44	8	11	3
2 words	5	14	12	7	5
3 words	3	2	8	3	9
4 words	4	0	11	1	13

To test the prediction that more illusory changes related to the verbal transformation effect (i.e., a different word or a non-word is reported) would be observed at shorter list lengths than at longer list lengths we conducted a 2 X 2 Chi Square test to compare the number of illusory words and non-words elicited for lists with 1 word to the number of illusory words and non-words elicited for lists with 4 words. The results showed that significantly more illusory words and nonwords were reported for lists with 1 word than for lists with 4 words ($\chi^2 = 15.795$, $df = 1$, $p < 0.0001$).

To test the prediction that more illusory changes related to the speech to song illusion (i.e., a change in pitch or rhythm is reported) would be observed at longer list lengths than at shorter list lengths we conducted a 2 X 2 Chi Square test to compare the number of changes in pitch or rhythm reported for lists with 1 word to the number of changes in pitch or rhythm reported for lists with 4 words. The results showed that significantly more changes in pitch or

rhythm were reported for lists with 1 word than for lists with 4 words ($\chi^2 = 7.615$, $df = 1$, $p = 0.0058$).

We did not have any *a priori* predictions regarding patternizations, and so did not perform any statistical analysis of them. Indeed, based on the previous literature, we did not expect to observe the class of illusory changes we have labeled as patternization. Therefore, we simply report the number of patternizations that occurred to demonstrate that this was not an anomalous observation nor an idiosyncratic percept. We also set aside this new category of percept for future research to examine further.

Discussion

In the present study, we combined methods typically employed to examine the verbal transformation effect with methods typically employed to examine the speech to song illusion. This resulted in participants hearing word lists varying in length (from 1 to 4 words) that were repeated for 4 minutes. In addition, rather than constrain the responses of participants to indicate if/when they experienced an illusion, or to rate the strength of the illusion, we employed a more open-ended approach to collecting responses from participants by simply asking them to indicate when an illusory change occurred and to describe the nature of the change.

Using a more open-ended approach to collecting responses from participants resulted in the discovery of a previously unreported percept, which we have called *patternization*. In patternization two versions of a given (non)word or (non)word list alternate with each other (e.g, "candle, cando, candle, cando…" when presented only with the word "candle"). This percept differs from the transformations typically observed in the verbal transformation effect, where the stimulus word is initially perceived veridically, but then is perceived as a different (non)word for some time, then shifts again to still another (non)word for some time, etc. Had we employed the

typical methods used to examine these auditory illusions, we would not have discovered this new illusory percept (which we presently set aside for future research to examine further).

Although the present study was somewhat exploratory in nature we did conduct analyses to determine whether the repetition account of the speech to song illusion (Deutsch et al., 2011; Margulis, 2013; Margulis & Simchy-Gross, 2016; Rowland et al., 2019) or the NST account of the verbal transformation effect (MacKay, Wulf, Yin & Abrams, 1993; Shoaf & Pitt, 2002) and the speech to song illusion (Castro et al., 2018; Mullin et al., 2021; Vitevitch et al., 2021) provided a better explanation for the percepts that were observed. We reasoned that if the repetition account brings about both the verbal transformation effect and the speech to song illusion, then the number of illusory changes reported should remain relatively constant over time. In contrast, if the NST account of the verbal transformation effect and the speech to song illusion brings about both illusions, then one would expect more illusory changes reported after 4 minutes of repetition than after 1 minute of repetition, because the satiation of nodes takes some time to occur, and many more nodes would be satiated later in the session than earlier in the session.

Comparison of the number of illusory changes reported in minute 1 to the number of illusory changes reported in minute 4 shows that significantly more illusory changes were reported later in the session (i.e., minute 4) than earlier in the session (i.e., minute 1). This finding is more consistent with the NST account of the verbal transformation effect and the speech to song illusion (where satiation of word nodes takes time to occur), than with the repetition account of the speech to song illusion. In the General Discussion we will compare in more detail various accounts of the verbal transformation effect and the speech to song illusion.

We also attempted to discern if the verbal transformation effect and the speech to song illusion are truly distinct illusions. If the two illusions are distinct, then distinct percepts should be reported for each list length (i.e., 1 to 4 words in the list). Recall that studies of the verbal transformation effect typically use 1 word as a stimulus, whereas studies of the speech to song illusion typically use phrases or word lists of about 4 words. Alternatively, if the illusions are not completely independent, the same percepts may be reported for shorter word lists and for longer word lists.

We compared the percepts often reported for the verbal transformation effect (i.e., words and non-words) for lists of length 1 and 4, and found that significantly more word and nonword percepts were reported for lists with 1 word than for lists with 4 words. This finding suggests that the verbal transformation effect and the speech to song illusion are somewhat distinct. However, the distinction between the two illusions is weakened by the fact that *any* word/nonword percepts were reported for lists of 4 words.

To further examine whether the verbal transformation effect and the speech to song illusion are distinct phenomena, we compared the percepts typically experienced in the speech to song illusion (i.e., changes to pitch and rhythm) for lists of length 1 and 4. If the two illusions are distinct, we expected that more percepts typically experienced in the speech to song illusion (i.e., changes to pitch and rhythm) would be reported for lists of length 4 (and few or no such reports for lists with 1 word). In this case, we found the opposite of what we predicted. That is, significantly more changes to pitch and rhythm percepts were reported for lists with 1 word than for lists with 4 words. Finding (significantly more) music-like percepts for lists with 1 word again weakens the distinction between the two illusions.

Note that in their study of the verbal transformation effect, Kaminska and Mayer (2002) also observed changes in music-like percepts (changes to pitch and emphasis/stress, but not to rhythm) in their word lists, which ranged from 1 to 2 words (and 1 to 4 syllables). Given the comparable nature of the stimuli in Kaminska and Mayer (2002) and in the present study, and the similarity in the findings of the two studies, observing "music-related" changes in lists with 1 word is unlikely to be the result of the stimuli or other methodological choices we made in the present study. Instead, the results from the study by Kaminska and Mayer (2002) on the verbal transformation effect and the results from the present study (ostensibly) on the speech to song illusion both suggest that the two illusions may not be as distinct as previous reports in the literature might lead one to believe.

The distinction in the literature between the verbal transformation effect and the speech to song illusion may have arisen in part due to the tasks employed in the previous studies, and the nature of the response that participants were able to provide in the previous studies. Had we not adopted a more open-ended approach to collecting participant responses in our study of the speech to song illusion, or had we repeated the stimuli for a shorter amount of time we might not have observed the blurring of the boundary between the verbal transformation effect and the speech to song illusion.

The distinction in the literature between the verbal transformation effect and the speech to song illusion may also have arisen due in part to the fact that researchers typically examine only one of these illusions at a time. It is important to note that in their study of the verbal transformation effect, Kaminska and Mayer (2002) found perceptual transformations for repeated non-speech stimuli, including music and "…other complex everyday sounds…" (e.g., coin dropping, car skidding). A number of the transformations they observed for music and

everyday sound stimuli were "music-related" changes, such as changes in pitch, rhythm, and rate. That is, the repetition of everyday, environmental sounds resulted in changes to the "music-like" percepts of the sounds. We were struck by the similarity between the finding of Kaminska and Mayer (2002) that the verbal transformation effect for environmental sounds could lead to music-like percepts, and the finding of Simchy-Gross and Margulis (2018) that the "speech to song" illusion could also be elicited by everyday environmental sounds (resulting in the sound to music illusion). The similarity between the verbal transformation effect and the speech to song/sound to music illusion with regards to environmental sounds in part motivated Study 2.

Study 2

In their study of the verbal transformation effect, Kaminska and Mayer (2002) found that participants reported music-related perceptual transformations for everyday sounds, such as a coin dropping, or a car skidding. That is, "music-related" transformations in pitch, rhythm, and rate were reported for everyday sounds that were repeated. A similar perceptual transformation was reported by Simchy-Gross and Margulis (2018) who found that the "speech to song illusion" could also be elicited by everyday environmental sounds that were repeated, a phenomenon they called the sound to music illusion. We were struck by the similarity between these two studies of different auditory illusions, and sought in the present study to examine if a stimulus that could be perceived as speech or could be perceived as (a non-speech) sound could also elicit a "music-like" illusory percept when repeated.

In order to create a stimulus that could be perceived as speech or could be perceived as a non-speech sound we used a vocoder to manipulate the acoustic signal of the word-lists with dense phonological neighborhoods used in Castro et al. (2018), which are known to elicit the speech to song illusion. Vocoders are often used to simulate for listeners with normal hearing

how sounds (e.g., speech, music, etc.) are heard by users of cochlear implants. It is important to note that there is some debate about how accurately a vocoder simulates sounds as experienced by a cochlear implant user (Dorman et al., 2020). For the present purposes, however, that issue is irrelevant. What is relevant is that speech transformed with a vocoder to simulate a cochlear implant with 1 channel sounds less like speech and results in poor speech recognition performance, whereas speech transformed with a vocoder to simulate a cochlear implant with more than 1 channel (*N.B.*, we used 12 channels in the present study) sounds increasingly like speech and results in increasingly better speech recognition performance (Dorman, Loizou, Fitzke & Tu, 1998).

In the present study we transformed lists of spoken words with a vocoder to simulate a cochlear implant with 1 channel to produce our noise-like stimuli. We transformed the same lists of spoken words with a vocoder to simulate a cochlear implant with 12 channels to produce our speech-like stimuli. We described to participants the speech to song illusion, and presented to them the phrase (repeated 10 times), "sometimes behave so strangely," (excised from Deutsch, 1995) to further illustrate the illusion. Participants then heard the vocoder-transformed word-lists (each list was repeated 10 times), and were asked to indicate *YES* or *NO* if each list elicited the speech to song illusion.

Based on the Node Structure Theory account of the speech to song illusion (Castro et al., 2018; Mullin et al., 2021; Vitevitch et al., 2021) we predicted that only stimuli perceived as speech would be susceptible to the speech to song illusion. We reasoned that only speech-like stimuli would activate word nodes that would satiate with repetition (resulting in the loss of the speech percept), leaving the syllable nodes to be repeatedly primed (resulting in the emergence of a rhythmic, song-like percept). Thus, only the word-lists that were vocoder-transformed to

simulate 12 channels of a cochlear implant (i.e., the speech-like stimuli) would evoke the speech to song illusion, whereas the word-lists that were vocoder-transformed to simulate 1 channel of a cochlear implant (i.e., the noise stimuli) would not evoke the speech to song illusion.

Methods

Participants: Sixty-two undergraduate students over the age of 18 years received participation credit for an Introductory Psychology class as compensation for their participation in this experiment. Participants were all native speakers of English, reported no speech or hearing disorders, and provided written informed consent. This study was initiated and completed prior to March 2020 (when various restrictions were imposed in the United States to limit the spread of the COVID-19 virus), resulting in a larger sample size compared to some of the other studies in the present report.

Materials: The 4-word lists of words with dense phonological neighborhoods used in Castro et al. (2018) were used in the present study. Pratt (Version 6.1.15; Boersma & Weenink, 1992) and a Pratt script (Vocoder; Version 47, August 2022 written by Matthew Winn http://www.mattwinn.com/praat.html) were used to transform the word-lists. The default settings for the parameters in the Vocoder script were used with the exception that (numberOfChannels) and (numberStimulated) were both set to 1 channel to produce our "noise" stimuli, and to 12 channels to produce our "speech" stimuli. We selected 12 channels in order to produce some distortion in the acoustic signal, but not so much distortion that speech recognition would be significantly impaired (Dorman et al., 1998).

Procedure: After providing written consent each participant heard an explanation of the speech to song illusion, and was presented with the phrase (repeated 10 times) "sometimes behave so strangely" (excised from Deutsch, 1995). Participants were then asked to indicate *YES* or *NO* if

they experienced the illusion or not. Participants then heard the vocoder-transformed word-lists (each list was repeated 10 times), and were asked to indicate *YES* or *NO* if each list elicited the speech to song illusion. The same equipment used in Study 1 was also used in the present study.

Results

Of the 62 participants initially recruited for the experiment, 5 indicated that they did not experience the speech to song illusion when presented with the phrase from Deutsch (1995). The remaining 57 participants responded that they did experience the speech to song illusion when presented with the phrase from Deutsch (1995). Only the data from the 57 participants who initially experienced the illusion with the phrase from Deutsch (1995) were analyzed further.

A two-tailed paired-samples *t*-test shows that the word-lists that were vocoder-transformed to simulate 12 channels of a cochlear implant (i.e., the speech-like stimuli) evoked the speech to song illusion more (*mean* = 40.1%; *sd* = 25.5) than the word-lists that were vocoder-transformed to simulate 1 channel of a cochlear implant (i.e., the noise stimuli; *mean* = 28.8%; *sd* = 23.9; t (56) = 2.6126, $p < 0.05$). We also used a one-sample (two-tailed) *t*-test to determine if the response rate for the word-lists that were vocoder-transformed to simulate 1 channel of a cochlear implant (i.e., the noise stimuli) differed significantly from a hypothetical mean response rate of 0%. The results show that the response rate to the "noise" stimuli differed significantly from zero (t (56) = 9.0976, $p < 0.0001$).

Discussion

In the present study we were motivated in part by prior observations that repetition of environmental sounds leads to illusory musical percepts, as observed by Kaminska and Mayer (2002) in a study of the verbal transformation effect, and by Simchy-Gross and Margulis (2018) in a study of the speech to song illusion. We transformed spoken words with a vocoder to create

speech-like and noise-like versions of the same stimulus, and asked participants to indicate (*YES* or *NO*) if the transformed stimuli evoked the speech to song illusion. Based on the NST account of the speech to song illusion (Castro et al., 2018; Mullin et al., 2021; Vitevitch et al., 2021), we reasoned that only speech-like stimuli would activate word nodes that would satiate with repetition (resulting in the loss of the speech percept), leaving the syllable nodes to be repeatedly primed (resulting in the emergence of a rhythmic, song-like percept). We therefore predicted that only the speech-like stimuli would evoke the speech to song illusion, whereas the noise stimuli would not evoke the speech to song illusion.

We found that the speech-like stimuli indeed evoked the speech to song illusion more than the noise-like stimuli, suggesting that a stimulus may need to be perceived first as speech for the speech to song illusion to occur. This finding is consistent with our prediction from the NST account of the speech to song illusion.

However, we also found that the noise stimuli evoked the speech to song illusion at a (significantly) non-zero rate. This was not as we predicted based on the NST account of the speech to song illusion. Evoking music-like percepts in the noise stimuli used in the present study is somewhat consistent with the findings of Kaminska and Mayer (2002) and of Simchy-Gross and Margulis (2018). Both of those previous studies found that participants reported music-related changes in their percepts of repeated words (Kaminska & Mayer, 2002) and in their percepts of repeated environmental sounds (Simchy-Gross & Margulis, 2018). Finding that the "noise" stimuli in the present study evoked the speech to song illusion at a non-zero rate was somewhat surprising, because the vocoder transformation to simulate a cochlear implant with 1 channel produces significant distortion to the acoustic signal. Typically, the resulting signal is described as *not* speech-like. Further, it is unlikely that the resulting signal would be categorized

as being a sound encountered in the natural environment (i.e., the stimuli used in Simchy-Gross & Margulis, 2018), making this a somewhat unexpected result.

Finding only partial support in the present study for the NST account of the speech to song illusion raised a number of questions. Some questions are methodological in nature, whereas others are theoretical in nature.

Regarding the methodological questions, in the present study we used the methodology typically associated with the speech to song illusion (i.e., 10 repetitions of the stimulus, only asking if the participant experienced the speech to song illusion). It is not clear whether different results would be obtained with use of the methodology typically associated with the verbal transformation effect (e.g., repetition of the stimulus for about 4 minutes), or with use of the more open-ended response format that we employed in Study 1. Given our long-standing interest in spoken word recognition (Siew & Vitevitch, 2016), speech production (Vitevitch et al., 2015), word-learning (Vitevitch et al., 2014), and bilingualism (Vitevitch, 2012), as well as our previous research on the speech to song illusion from the perspective of a language processing model (Castro et al., 2018; Mullin et al., 2021; Vitevitch et al., 2021), we will continue in the remaining studies reported here to examine various aspects of the speech to song illusion and the potentially related sound to music illusion. We leave it to future research to answer this broader set of methodological questions.

Turning to the theoretical questions, is there a single underlying mechanism that is responsible for the all of the auditory illusions that have been discussed in the present set of studies, namely, the verbal transformation effect to other speech sounds (i.e., words or nonwords), the verbal transformation effect to non-speech sounds (i.e., changes in pitch or rhythm), the speech to song illusion, and the sound to music illusion? If there is a single

underlying mechanism that is responsible for the all of these auditory illusions, then which theoretical approach provides the better account of that single underlying mechanism—NST, the repetition account, or some other theoretical approach? If there is more than one mechanism that underlies all of these auditory illusions, then what is the nature of those mechanisms? Are they related to language-processing, music-processing, or to general auditory processing? Are the mechanisms cognitive/"higher-level" in nature or perceptual/"lower-level" in nature?

In addition to these methodological and theoretical questions, there are additional questions related to how the single mechanism or multiple mechanisms may be instantiated neurologically and physiologically. It is unlikely that a single study will answer all of these questions. It is also unlikely that the remaining studies in the present report will completely answer any of these questions. Nevertheless, the data that we present in the present studies on the speech to song illusion and the potentially related sound to music illusion will provide some insight in to some of these questions, and guide future research.

Study 3

In this study we narrowed our focus just to the speech to song illusion, and attempted to address the question of whether the mechanism that underlies the speech to song illusion is cognitive/"higher-level" in nature or perceptual/"lower-level" in nature. The evidence we provide in this study might also help distinguish between the Node Structure Theory and the repetition accounts of the speech to song illusion.

Because much research has examined how lower-level, acoustic parameters of the stimulus influence the speech to song illusion (e.g., Falk, Rathcke & Dalla Bella, 2014), we decided in the present study to focus on a cognitive/"higher-level" characteristic of the stimulus. We reasoned that a mechanism that was cognitive/"higher-level" in nature would be influenced

by manipulations to cognitive/"higher-level" characteristics of the stimulus, whereas a mechanism that was perceptual/"lower-level" in nature would not be influenced by manipulations to cognitive/"higher-level" characteristics of the stimulus.

The cognitive/"higher-level" characteristic of the stimulus that we chose to manipulate in the present study was phonological neighborhood density, which refers to the number of words in that part of memory known as the mental lexical that sound similar to the stimulus word (Vitevitch & Luce, 2016). Phonological neighborhood density was manipulated in Experiment 1 in Castro et al. (2018), where they found that lists containing words with many similar sounding words (i.e., a dense phonological neighborhood) were rated as being more song-like in a study of the speech to song illusion than lists containing words with few similar sounding words (i.e., a sparse phonological neighborhood). In the present study, we used a different methodology to test whether phonological neighborhood density influenced some other aspect of the speech to song illusion, namely the point in time at which the percept changes from speech to song. Typically studies of the speech to song illusion ask participants to rate on a Likert scale the speech- or song-likeness of the stimulus to assess the strength of the illusion, or simply indicate (yes/no) if they experienced the illusion (as in Study 2). Note that all three approaches—time to transform, yes/no, Likert scale—were employed in Mullin et al. (2021), demonstrating the utility of each method to investigate the speech to song illusion.

To determine when the percept changes from speech to song we employed in the present study a task that had been previously employed in Study 3 of Mullin et al. (2021). Participants listened to a list of words repeated multiple times and were instructed to click a button labeled "I experienced the illusion" upon experiencing the perceptual shift. If they did not experience the illusion, they were instructed to click the button labeled "I did NOT experience the illusion" after

hearing the stimulus. We used the time at which the "I experienced the illusion" button was clicked as a measure of when the percept changed from speech to song.

If the mechanism underlying the speech to song illusion was cognitive/"higher-level" in nature, we predicted that there would be a difference in when the percept changed from speech to song for the lists of words varying in a cognitive/"higher-level" variable, namely phonological neighborhood density. If the mechanism underlying the speech to song illusion was perceptual/"lower-level" in nature, we predicted that there would be no difference in the point in time at which the percept changed from speech to song for the lists of words that varied in phonological neighborhood density.

We further predicted that if Node Structure Theory was the mechanism that better explains the speech to song illusion, then lists containing words with dense phonological neighborhoods (i.e., many similar sounding words) would elicit a perceptual shift from speech to song earlier than lists containing words with sparse phonological neighborhoods (i.e., few similar sounding words). Recall that Castro et al. (2018) found that lists containing words with dense phonological neighborhoods were rated as being more song-like than lists containing words with sparse phonological neighborhoods. They suggested that the amount of priming transmitted by phonological nodes to the lexical nodes varied for the dense and sparse words, with more priming being transmitted to sparse words, allowing them to recover from satiation more quickly than dense words, and therefore decreasing the song-like percept (and ratings) for sparse words. The difference in the amount of priming transmitted by phonological nodes to the lexical nodes for dense compared to sparse words would also result in the dense words shifting from speech to song earlier than the sparse words.

As discussed in Castro et al. (2018), it is unclear how the repetition account could explain the difference in song-like ratings for the dense and sparse word-lists. Similarly, the repetition account would not be able to explain a difference in when the perceptual shift occurs should it be observed for the dense and sparse word-lists in the present study.

Thus, the previously employed methodology of asking when the percept shifted from speech to song (Mullin et al., 2021) not only allowed us to adjust to the changes to our research program that were required due to COVID-related policies, but also enabled us to provide some evidence that would address some of the theoretical questions described above. Specifically, if the mechanism that underlies the speech to song illusion is cognitive/"higher-level" in nature, then a difference should be observed in when the percept shifts from speech to song for the word list varying in phonological neighborhood density. If no difference is observed, then the possibility that a perceptual/"lower-level" mechanism underlies the speech to song illusion remains viable.

Further, if the percept shifts earlier for dense word-lists than sparse word-lists, that finding would be consistent with the predictions derived from the NST account of the speech to song illusion. However, if no difference is observed, then that finding would be a challenge to the NST account of the speech to song illusion.

Methods

The current study was initiated shortly after March 2020, which is when various restrictions were imposed in the United States to limit the spread of the COVID-19 virus. These restrictions required us to use remote technologies to collect data. The time at which the present study was initiated during the semester also limited the number of participants that could be recruited during that academic semester.

Participants: One-hundred eighty-six undergraduate students over the age of 18 years received participation credit for an Introductory Psychology class as compensation for their participation in this on-line experiment. Participants were all native speakers of English, reported no speech or hearing disorders, and provided written informed consent.

Materials: The study was administered in English through an internet-based Qualtrics survey. A captcha was presented at the beginning of the survey as an initial screening for bots (i.e., non-human robots that automatically respond to survey questions, or aid humans to respond more rapidly to survey questions; Kennedy et al., 2020).

Two lists containing words with dense phonological neighborhoods ([1] cattle banner tackle hurry; [3] dairy meter body lighter) and two lists containing words with sparse phonological neighborhoods ([2] cashew burden tower hero; [4] devil mighty bottom lotion) that were previously used in Castro et al. (2018) were used in the present study. Each list was repeated 10 times. The number in square brackets next to each wordlist above indicates the order of presentation in the Qualtrics survey.

Procedure: Once written e-consent was received, participants were presented with the following instructions:

> *Illusions occur when we incorrectly perceive what is in our environment. In this study, we are interested in an auditory illusion called the speech to song illusion. In the speech to song illusion, a spoken phrase is repeated multiple times. After several repetitions, the spoken phrase is heard by some listeners as more "song-like" rather than being spoken. We want to know when you experience this perceptual shift. In this study, you will listen to 4 audio clips of a spoken phrase. Each phrase will repeat several times. If the phrase shifts from speech to song please click the button labeled "I experienced the illusion" as soon as the shift occurs. If you do not experience this illusion (not all people do) then simply wait until the sound file is done playing and then click the button labeled "I did NOT experience the illusion." The next sound file will then be presented.*

The sound files were programmed to start playing automatically, allowing us to track the amount of time that elapsed until the button labeled "I experienced the illusion" had been

clicked. Because the default settings of mobile devices (i.e., phones, tablets, etc.) prevents audio and video media from starting automatically (in order to prevent undesired charges to data plans, etc.) participants were instructed when signing up for the study to only participate using a laptop or desktop, not a phone or tablet. Despite that warning, a number of participants still attempted to participate in the experiment using a phone or tablet, resulting in technical difficulties (i.e., the sound files would not play). As noted below the data from these participants were not included in the analyses.

At the end of each sound file participants were asked: *What 4 words were being repeated in the audio file?* Data from participants who did not respond correctly were excluded from the analysis. After the participant answered this question, the next audio file began to play automatically.

Results

One-hundred eighty-six participants started the Qualtrics survey, but only 129 completed the survey. Of those that completed the survey, 11 indicated a technical problem (i.e., "no audio") when asked to respond to the question: *What 4 words were being repeated in the audio file?* The information provided by Qualtrics regarding the browser and operating system of those participants confirmed that a mobile device instead of a laptop or desktop computer was used to complete the survey. Finally, 1 participant did not provide any correct responses to the question: *What 4 words were being repeated in the audio file?* The data from this participant and those participants experiencing technical problems were not included in the analyses.

The time that elapsed from the start of the sound file (which started playing automatically) until the button labeled "I experienced the illusion" was clicked served as the dependent variable. The time that elapsed until the button labeled "I did NOT experience the

illusion" was clicked was not included in the analyses. Because Mullin et al. (2021) found that approximately 3-5 repetitions of a phrase/list were needed to elicit the speech to song illusion, we excluded "I experienced the illusion" responses if they were less than 2 seconds (i.e., less than 1 repetition of the word list had occurred; 5 trials). Because the sound files were approximately 20 seconds in duration we also excluded "I experienced the illusion" responses if they were more than 10 seconds (i.e., the participant did not "click the button labeled "I experienced the illusion" as soon as the shift occurs." [emphasis added]; 124 trials).

Because there was some variability in experiencing the illusion (i.e., some participants reported experiencing the illusion for all 4 lists, some did not) we considered each response to be independent, and used an independent samples one-tailed *t*-test (i.e., we predicted that *dense* should be less than *sparse*) in JASP (2022). We found that participants responded to the lists of words with dense phonological neighborhoods more quickly (*mean* = 5.76 seconds; *sd* = 2.01; *n* = 74 responses) than to lists of words with sparse phonological neighborhoods (*mean* = 6.41 seconds; *sd* = 2.02; *n* = 50 responses; t (122) = -1.76, $p < .05$).

Discussion

In the present study we examined when the speech to song illusion occurred by measuring the time at which the percept shifted from speech to song for lists of words varying in phonological neighborhood density that were repeatedly presented to participants. We used a method that had been previously employed to examine the speech to song illusion in younger and older adults (Mullin et al, 2021), and found that the lists of words with dense phonological neighborhoods (i.e., the words sounded similar to many other words) evoked the speech to song illusion more quickly than lists of words with sparse phonological neighborhoods (i.e., the words sounded similar to few other words). This finding is consistent with the findings from Castro et

al. (2018), who used a 5-point rating scale of song-likeness and found that the lists of words with dense phonological neighborhoods were rated as more song-like than lists of words with sparse phonological neighborhoods. This finding is also consistent with our predictions derived from Node Structure Theory.

Observing that a cognitive/"higher-level" variable, namely phonological neighborhood density, influenced when the speech to song illusion occurred provides some support to the idea that the mechanism underlying the speech to song illusion is cognitive/"higher-level" in nature. To be clear, this observation does not definitively rule out the possibility that there could be another mechanism involved in the speech to song illusion, nor that the additional mechanism may be perceptual/"lower-level" in nature.

It is not clear how the repetition account of the speech to song illusion would explain the variation in the onset time of the perceptual shift between phonological neighborhood density conditions. The repetition account would predict no difference between these conditions. That a significant difference was observed calls into question the repetition account as the mechanism that underlies the speech to song illusion.

Given our interest in the speech to song illusion, and our interest in the mechanism(s) that underlies all of the auditory illusions we have discussed, we examined in the remaining studies when the sound to music illusion occurs. We focused on the sound to music illusion (Simchy-Gross & Margulis, 2018), in which there is a perceptual shift from environmental sounds to a music-like percept, because it more closely parallels the speech to song illusion. In contrast, the perception of music-like transformations to a repeated stimulus observed by Kaminska and Mayer (2002) and in Study 1 more closely parallels the verbal transformation effect for words.

Study 4

In this study we explored the possibility that the *music-related illusions*—the speech to song illusion and the sound to music illusion—may have a common underlying mechanism. We used the methodology that was used to examine the speech to song illusion in Study 3 and in Mullin et al. (2021) to now examine the sound to music illusion (Simchy-Gross & Margulis, 2018). Specifically, participants heard sound files that contained environmental sounds (water dripping, shovel scrapping gravel, ice cracking, whale song) that were repeated for approximately 20 seconds. Participants were instructed to click a virtual button upon experiencing the perceptual shift from sound to music, allowing us to measure the time at which the perceptual shift occurred.

If the speech to song illusion and the sound to music illusion are produced by different mechanisms, we reasoned that when the percept shifted from speech to song (as determined in Study 3) would be different from when the percept shifted from sound to music. Observing a difference in the timing of the perceptual shift between the two illusions would be strong, but not definitive, evidence that more than one mechanism may be involved in producing the speech to song illusion and the sound to music illusion. We state that this evidence would not be definitive because this finding could not rule out the possibility that a common mechanism underlies both illusions with certain percepts being more likely to occur earlier versus later in the repetition of the signal (as observed in Study 1).

Similarly, if there is no difference in when the percept shifts from speech to song and from sound to music, one could interpret that as evidence for a common mechanism underlying both illusions. However, that finding would not be able to rule out the possibility that the illusions are produced by different mechanisms that coincidentally follow the same time-course,

perhaps due to the fundamentals involved in processing any kind of auditory signal, or due to the dynamics of auditory signals in general. Although neither result would provide definitive evidence to support any of the possibilities we have outlined above, the results we obtain from the present study will provide guidance to future researchers probing the perceptual system in studies of these auditory illusions.

Methods

The present study was initiated after March 2020 (when various restrictions were imposed in the United States to limit the spread of the COVID-19 virus), which required us to use remote technologies to collect data. This study was also initiated at the beginning of an academic semester, enabling us to recruit throughout the whole semester to obtain a large sample of participants.

Participants: Three-hundred fifty-one undergraduate students over the age of 18 years received participation credit for an Introductory Psychology class as compensation for their participation in this on-line experiment. Because this was not a language-based illusion we did not restrict the participants in the present study to being native speakers of English. None of the participants reported a speech or hearing disorder, and all provided written informed consent.

Materials: A Qualtrics survey in English (and captcha) similar to the one used in Study 3 was used in the present study. We selected four environmental sounds—water dripping, a shovel scraping gravel, ice cracking, and whale song—to use as stimuli for this study. These sounds were selected based on Simchy-Gross & Margulis (2018), who reported that these four environmental sounds from among the sounds they tested exceeded a rating of 3 on a 5-point Likert scale rating perception of song-likeness. We acknowledge that Kansas is a doubly-landlocked state (meaning that a minimum of two states must be traversed to access an ocean,

bay, or gulf), which raises the concern that the typical undergraduate student enrolled at the University of Kansas is unlikely to frequently encounter whale song in their everyday environment. We set aside that concern, and obtained from various free sound effects websites sound files of those 4 sounds to use in the present study. We used Audacity 2.3.3 to excise a segment of each sound file that was approximately 2 seconds in duration. Each sound file was reduplicate to create sound files approximately 20 seconds in duration.

Procedure: Once written e-consent was received, participants were presented with the following instructions:

> *Illusions occur when we incorrectly perceive what is in our environment. In this study, we are interested in an auditory illusion called the sound to song illusion. In the sound to song Illusion, a sound commonly found in the environment is repeated multiple times. After several repetitions, the sound is heard by some listeners as being more "music-like." We want to know when you experience this perceptual shift. In this study, you will listen to 4 audio clips of sounds commonly found in the environment. Those sounds will repeat several times. If the sound becomes more music-like please click the button labeled "I experienced the illusion" as soon as the shift occurs. If you do not experience this illusion (not all people do) then simply wait until the sound file is done playing and then click the button labeled "I did NOT experience the illusion."*

As in Study 3, the sound files were programmed to start playing automatically, allowing us to track the amount of time that elapsed until the button labeled "I experienced the illusion" had been clicked. Again, as in Study 3, participants were instructed to complete the experiment using a laptop or desktop computer rather than a cell phone or tablet, although a number of participants still attempted to complete the survey on inappropriate devices. The data from these participants were not included in the analyses.

At the end of each sound file participants were asked: *What sound do you think was being repeated in the audio file?* and were presented with the following options: *water dripping, whale song, ice cracking, shovel dragging, Did not hear a sound/technical problem, The sound I heard*

was none of the above. The data from participants were not included in the analyses if they did not respond to the question correctly. After answering this question, the next audio file would begin to play automatically.

Results

Three-hundred fifty-one participants started the Qualtrics survey, but only 295 participants completed the survey. The time that elapsed from the start of the sound file (which started playing automatically) until the button labeled "I experienced the illusion" was clicked served as the dependent variable. The time that elapsed until the button labeled "I did NOT experience the illusion" was clicked was not included in the analyses. In order to better compare the results of the present study to the results of Study 3 we used the same cut-offs that were used in Study 3. That is, responses less than 2 seconds and greater than 10 seconds were excluded from the analysis (resulting in the loss of 273 responses).

As in Study 3, there was some variability in experiencing the illusion (i.e., some participants reported experiencing the illusion for all 4 sounds, some did not), so we again considered each response to be independent, and conducted an independent samples ANOVA in JASP (2022) to compare the response times between conditions. There was not a significant difference among the 4 sound files with regard to the onset of experiencing the perceptual shift (F (3, 261) = 1.52, p = .21). See Table 4 for the summary data for each sound file.

Table 4. Summary data for the four environment sounds used in Study 4 to examine the sound to music illusion.

Sound	Mean Click-time (seconds)	sd	n
Ice cracking	5.80	2.05	22
Shovel on gravel	5.87	2.36	50
Water dripping	5.65	2.17	100
Whale song	6.34	2.38	93

We used a one-sample *t*-test to compare the time at which the perceptual shift occurred in Study 3 to the time at which the perceptual shift occurred for the environmental sounds obtained in the present study. We used as the hypothetical mean the mean time to click of all four environmental sounds (*mean* = 5.92). The mean of the dense and sparse conditions and the mean of the standard deviations for those conditions were used as the actual mean (*mean* = 6.09) and standard deviation (*sd* = 2.01). A two-tailed one sample *t*-test showed that there was no difference between the time to click for speech and the time to click for environmental sounds ($t(123) = 0.9465$, $p = 0.35$).

Discussion

In the present study we investigated the time course of the perceptual shift experienced in the sound to music illusion using the methodology previously employed in Study 3 and in Mullin et al. (2021) with the aim of furthering our understanding of the speech to song illusion, and determining if the two auditory illusions are indeed distinct phenomena. By comparing the time at which the sound to music illusion emerged to the time at which the speech to song illusion emerged, we also hoped to provide evidence regarding the mechanisms underlying the two illusions. We reasoned that two different timeframes for the two illusions would implicate two

different mechanisms involved in the illusions (or perhaps of one mechanism with certain percepts emerging earlier with repetition and other percepts emerging after more repetition as was observed in Study 1). However, the results of the present study show that there was no difference in when the two illusions emerged, which blurs the distinction between the two auditory illusions.

Now consider whether there is a common mechanism that underlies these two music-related illusions. One interpretation of the similar time courses of these illusions is that there is indeed a common mechanism underlying both illusions. In addition to being based on a null-result that conclusion is problematic because the present data cannot rule out the possibility that the illusions are produced by different mechanisms that coincidentally follow the same time-course. Therefore, we tried one last time in Study 5 to distinguish in some way the speech to song illusion from the sound to music illusion.

Study 5

In Studies 1-4 the speech and non-speech sounds used as stimuli were "neutral" or unlikely to evoke an emotional response. However, there are many speech and non-speech sounds that do evoke an emotional response. Consider, for example, work on phonaesthetics, which suggests that the sounds found in certain languages are perceived as more pleasant sounding than the sounds found in other languages (e.g., Winkler, Kogan & Reiterer, 2023). There is also much research showing that music can produce strong (positive or negative) emotional responses (e.g., Arjmand, Hohagen, Paton & Rickard, 2017). The emotional responses to music can be so strong that they may also influence how one experiences the taste of food in a phenomenon known as "sonic seasoning" (e.g., Xu, Guo, Liu, Xu & Huang, 2023). Given the musical nature of the speech to song illusion and the sound to music illusion, we sought in the

present study to exploit the emotional response experienced with speech and non-speech sounds, especially music, in an attempt to distinguish the two illusions from each other.

Previous work on the speech to song illusion examined how the use of word lists composed of words varying in emotional arousal might influence the illusion (Vitevitch et al., 2021). As predicted by Node Structure Theory, Vitevitch et al. (2021) found that phonological characteristics of the words influenced song-like ratings, but the emotional arousal of the words used in the word lists did not influence song-like ratings. Node Structure Theory predicts that once a word node is satiated the semantic information (such as emotional arousal) that is associated with the word node is no longer available to conscious awareness, and therefore could not affect song-likeness ratings. Given the absence of an "emotional" effect in the speech to song illusion, we wondered how emotion might influence the sound to music illusion. If the speech to song illusion is not influenced by emotion, but the sound to music illusion is influenced by emotion in some way, then that would distinguish the two illusions from each other, and might also suggest distinct mechanisms underlying the two illusions.

We considered a few options for emotional sounds or sounds that evoke a sense of "biological urgency" (Franconeri & Simsons, 2003), including sounds that induce the looming bias (McGuire, Gillath & Vitevitch, 2016), and screams (Arnal, Flinker, Kleinschmidt, Giraud & Poeppel, 2015). We were unable to devise a way to use the former stimulus option in paradigms commonly used to test the sound to music illusion, and the latter stimulus option seemed unlikely to be approved by the Institutional Review Board at the University of Kansas.

We then considered music, which is well-known to elicit intense emotional and psychophysiological responses, known as frission (Harrison & Loui, 2014). However, starting out with music as the sound stimulus makes it difficult to transform the percept from sound to

music to test the sound to music illusion. Fortunately, non-musical sounds have also been shown to induce frission (Honda et al., 2020). Therefore, we used one of the frission-inducing non-musical stimuli from Honda et al. (2020) to test the sound to music illusion using the time to click paradigm used in Studies 3-4.

Methods

The present study was initiated after March 2020, which is when various restrictions were imposed in the United States to limit the spread of the COVID-19 virus, and which forced us to use remote technologies to collect data. This on-line study was initiated in the middle of an academic semester, which limited the size of the sample that could be recruited.

Participants: One-hundred forty-one undergraduate students over the age of 18 years received participation credit for an Introductory Psychology class as compensation for their participation in this on-line experiment. Because this was not a language-based illusion, we did not restrict the participants in the present study to being native speakers of English. None of the participants reported a speech or hearing disorder, and all provided written informed consent.

Materials: A Qualtrics survey in English (and captcha) similar to the one used in Studies 2-4 was used in the present study. We downloaded the three Supplemental Audio files (of rolling glass beads) from Honda et al. (2020) to use as stimuli in the survey. Each file was 30 seconds in duration, so we used the files in their entirety without any additional modification.

Procedure: Once written e-consent was received, participants were presented with the following instructions:

> *Illusions occur when we incorrectly perceive what is in our environment. In this study, we are interested in an auditory illusion called the sound to song Illusion. In the sound to song Illusion, a sound commonly found in the environment is repeated multiple times. After several repetitions, the sound is heard by some listeners as being more "music-like." We want to know when you experience this perceptual shift. In this study, you will listen to 3 audio clips of sounds commonly found in the environment. Those sounds will*

repeat several times. If the sound becomes more music-like please click the button labeled "I experienced the illusion" as soon as the shift occurs. If you do not experience this illusion (not all people do) then simply wait until the sound file is done playing and then click the button labeled "I did NOT experience the illusion."

After participants indicated if they experienced the illusion or not, they were asked the following question: *Did the sound in the audio file give you a feeling of coldness or shivering in the absence of a physically cold stimulus?* This question was derived from the definition of frission in Honda et al. (2020). A “yes” response would indicate that the participant experienced frission, whereas a “no” response would indicate that the participant did not experience frission. After answering this question, the next audio file would begin to play automatically.

As in the previous studies, the sound files were programmed to start playing automatically, allowing us to track the amount of time that elapsed until the button labeled "I experienced the illusion" had been clicked. Again, as in the previous studies, a number of participants ignored the instruction to only participate in the experiment using a laptop or desktop, not a phone or tablet, resulting in technical difficulties (i.e., the sound files would not play) for those individuals. The data from these participants were not included in the analyses.

Results

One-hundred forty-one participants started the Qualtrics survey, but only 129 participants completed the survey. The time that elapsed from the start of the sound file (which started playing automatically) until the button labeled "I experienced the illusion" was clicked served as the dependent variable. The time that elapsed until the button labeled "I did NOT experience the illusion" was clicked was not included in the analyses. In order to better compare the time to experience the illusion obtained in the present study to the time to experience the illusion obtained in the previous studies in this report we used the same cut-offs that were used in the

previous studies, namely responses less than 2 seconds and greater than 10 seconds were excluded from the analysis (resulting in the loss of 128 responses).

Because of the variability in experiencing the illusion (i.e., some participants reported experiencing the illusion for all 3 sounds, some did not), we again considered each response to be independent, and used an independent samples ANOVA in JASP (2022) to analyze the response times. For the responses in which the sound to music illusion was experienced, we considered the time to click for when frission was experienced separately from when frission was not experienced.

There was not a significant difference among the 3 sound files for the time to experience the sound to music illusion (F (2, 69) = 1.15, p = .32), nor was there a significant difference for the time to experience the sound to music illusion between trials in which frission was experienced and trials in which it was not (F (1, 69) = 0.009, p = .92). Although the interaction of sound files and frission was statistically significant (F (2, 69) = 19.84, p = .04), none of the post-hoc comparisons were statistically significant when subjected to Tukey corrections for multiple comparisons. See Table 5 for the summary data for each sound file.

Table 5. Summary data for the 3 frission-inducing sounds used in Study 5 to examine the sound to music illusion.

List	Frission	Mean Click time (seconds)	sd	n
1	No Yes	5.96 7.20	2.68 1.67	9 10
2	No Yes	6.48 7.23	2.73 3.05	22 8
3	No Yes	6.93 4.76	1.88 1.99	15 11

Discussion

In the present study, we considered how "emotional" stimuli might distinguish the speech to song illusion from the sound to music illusion. A previous study of the speech to song illusion found that phonological characteristics of the words influenced song-like ratings, but the emotional arousal of the words used in the word lists did not influence song-like ratings (Vitevitch et al., 2021). The differential influence of phonological but not semantic/emotional information was consistent with predictions derived from the Node Structure Theory account of the speech to song illusion.

To examine how "emotion" might influence the sound to music illusion, we used a non-musical environmental sound (from Honda et al., 2020) that induces the intense emotional and psychophysiological response known as frission (Harrison & Loui, 2014). Given the absence of an "emotional" effect in the speech to song illusion, we reasoned that observing an influence of emotion in the sound to music illusion would suggest that these auditory illusions are distinct in some way.

Participants in the present study heard repetitions of a non-musical environmental sound (from Honda et al., 2020) that induced frission, and were asked to indicate when they experienced a shift in the percept from sound to music. The results showed that the time to experience the perceptual shift was consistent across different frission-inducing stimuli, and also across stimuli for which frission was and was not experienced. Given the previous reports of environmental sounds inducing the sound to music illusion, we expected stimuli that did not induce frission to cause a perceptual shift from sound to music (as was observed in the present study). The fact that frission-inducing stimuli caused a shift in the percept from sound to music at all is suggestive that the sound to music illusion differs in some way from the speech to song illusion, where "emotional" stimuli did not differentially affect song-likeness ratings.

The different way that "emotional" stimuli influence the speech to song illusion and the sound to music illusion raises questions about the mechanism that underlies these two illusions. Specifically, is there a single and common mechanism that produces the two illusions, or are there two (or more) mechanisms involved in what appears to be—based on the findings from the present study—two different illusions? In the General Discussion we will consider several accounts of the speech to song illusion, whether any of them can account for the findings from the present set of studies, and how many mechanisms might be needed to fully account for all of the auditory illusions examined in the present set of studies.

General Discussion

In five studies using a variety of tasks and stimuli we examined language-based auditory illusions (i.e., verbal transformation effect and the speech to song illusion) and music-based auditory illusions (i.e., the speech to song illusion and the sound to music illusion). In Study 1, methods typically used to study the verbal transformation effect were combined with methods

typically used to study the speech to song illusion. That is, word lists varying in length (from 1 word as is often used in the verbal transformation effect to 4 words as is often used in the speech to song illusion) were repeated for 4 minutes as is often done in studies of the verbal transformation effect (instead of the 10 repetitions often used in studies of the speech to song illusion). Further, rather than constrain participant responses (e.g., time to transform, yes/no, Likert scale; Mullin et al., 2021) we asked participants to report any and all changes to the percept that they experienced.

The results of Study 1 revealed an illusory percept that has not, to our knowledge, been reported previously that we call "patternization." In patternization two versions of a given word or word list alternate with each other. For example, the word "candle" might alternate with the nonword "cando," becoming "candle, cando, candle, cando…" This percept differs from what is often reported in the verbal transformation effect where the percept of a single word, like *flame*, changes to a series of other words, such as *blame*, then to *lame,* then to *fame,* etc. (Warren & Gregory, 1958).

We also found in Study 1 that more perceptual changes were reported later in the session than earlier in the session, and that the types of perceptual changes that were reported were not as unique to the verbal transformation effect or to the speech to song illusion as the previous literature on these two auditory illusions might lead one to believe. Specifically, music-related changes to the percept occurred at all list lengths, not just at the longer list lengths that resemble the stimuli used in the speech to song illusion (see Kaminska & Mayer (2002) for reports of musical percepts in their study of the verbal transformation effect). Language-related changes to the percept (e.g., different words or nonwords) also occurred at all list lengths, not just at the shorter list lengths that resemble the stimuli used in the verbal transformation effect. Previous

studies of these two auditory illusions may have made them appear more distinct than the data from Study 1 suggests due to the variation in the tasks that are typically used to examine each of them, due to the specific responses that participants were constrained to report in prior investigations of each illusion, or due to the fact that the illusions are often examined independently. The results of Study 1 suggest that a wider range of methods and response options might be needed to better understand the auditory illusions that we examined here. Further, combinations of methods, as was done in Study 1, might be needed to better examine the conditions under which various illusions occur.

Study 2 attempted to distinguish the speech to song illusion from the sound to music illusion by using a stimulus that sounded more noise-like or more speech-like. A vocoder was used to transform the same lists of spoken words to be more noise-like (i.e., a vocoder simulating a cochlear implant with 1 channel) or more speech-like (i.e., a vocoder simulating a cochlear implant with 12 channels). Participants were then asked to indicate *YES* or *NO* if each list elicited the speech to song illusion. We predicted that only the speech-like stimuli would elicit the speech to song illusion. The speech-like stimuli did indeed receive more YES responses than the noise stimuli. However, the noise-like stimuli also received a significant number of YES responses. Just as Study 1 showed that the verbal transformation effect is not as distinct from the speech to song illusion as previously thought, the YES responses to the noise-like stimuli in Study 2 suggest that the speech to song illusion may not be as distinct from the sound to music illusion as previously thought (Simchy-Gross & Margulis, 2018).

Studies 3-5 used a different experimental paradigm that asked participants to indicate the point in time that the percept shifted from speech to song to examine when the speech to song illusion occurs. In Study 3 we attempted to determine if the mechanism that produces the speech

to song illusion was cognitive/"higher-level" or perceptual/"lower-level" in nature. We reasoned that manipulations to cognitive/"higher-level" characteristics of the stimulus would only affect the speech to song illusion if the mechanism underlying it was cognitive/"higher-level" in nature. To that end we manipulated the psycholinguistic variable known as phonological neighborhood density (Vitevitch & Luce, 2016) and repeatedly presented lists of words that had either dense or sparse phonological neighborhoods. Participants' responses indicated that the percept changed from speech to song more quickly for the lists of dense words than for the lists of sparse words. This result suggests that the mechanism underlying the speech to song illusion is cognitive/"higher-level" in nature, and is consistent with the Node Structure Theory account of the speech to song illusion.

Given the blurring of the distinction between the speech to song illusion and the sound to music illusion suggested by the results of Study 2, we examined in Studies 4 and 5 when the sound to music illusion occurs using the "time to click" paradigm employed in Study 3. By using the same time to click paradigm employed in the investigation of the speech to song illusion in Study 3 we sought to determine if the sound to music illusion had a different time-course than the speech to song illusion. Observing a different time course for the two illusions would suggest that different mechanisms might underlie the two illusions. Using 4 environmental sounds, we found that the time course of the sound to music illusion is comparable to the time course of the speech to song illusion, making it difficult to definitively conclude if a common mechanism or different mechanisms underlie the speech to song illusion and the sound to music illusion.

In a final attempt to distinguish the speech to song illusion from the sound to music illusion, we used "emotional" stimuli to elicit the sound to music illusion. A previous study of the speech to song illusion found that words varying in emotional arousal did not differentially

affect song-like ratings in a speech to song illusion task (Vitevitch et al., 2021). In the present study of the sound to music illusion we used non-musical sounds (from Honda et al., 2020) that evoke the intense emotional and psychophysiological response known as frission (Harrison & Loui, 2014). In contrast to the lack of an influence of emotion in the speech to song illusion (Vitevitch et al., 2021), the frission-inducing sounds elicited the sound to music illusion. Further, the time course of the perceptual shift was similar to that observed in Study 3 for the speech to song illusion, again making it difficult to definitively conclude if a common mechanism or different mechanisms underlie the speech to song illusion and the sound to music illusion. In the next section we consider how the repetition account (Deutsch et al., 2011; Margulis, 2013; Margulis & Simchy-Gross, 2016; Rowland et al., 2019) and Node Structure Theory account (Castro et al., 2018; MacKay et al., 1993) may explain the findings from the present set of studies.

Previous accounts of the speech to song illusion

One account of the speech to song illusion suggests that repetition of the stimulus inhibits parts of the brain involved in language processing allowing the parts of the brain involved in music processing to become more active, resulting in a music-like percept (Deutsch et al., 2011; Margulis, 2013; Margulis & Simchy-Gross, 2016; Rowland et al., 2019). We have described above in the Introduction how this account is inconsistent with what is known about language processing. At the neuropsychological level, it is unclear how repetition of a stimulus *activates* the neural circuitry involved in the speech to song illusion when typically, repetition of a stimulus acts to *habituate* neural circuitry (e.g., Thompson & Spencer, 1966). Finally, at the cognitive level repetition of a stimulus seems to have different effects than what is suggested in the repetition account of the speech to song illusion. Consider the "mere exposure effect"

reported by Zajonc (1968), which suggests that repeated exposure to a stimulus increases processing fluency. If repeated exposure to a stimulus increases processing fluency, then repetitions of words/phrases should increase processing fluency of the words/phrases in the stimulus making the word/phrase more speech-like with repetition, not more music-like as is observed in the speech to song illusion. The different ways that repetition influences a stimulus in the "mere exposure effect" and in the speech to song illusion further undermines the repetition account.

Setting aside all of these concerns, if repetition does somehow enhance the musicality of a speech or non-speech stimulus as suggested by the repetition account (Deutsch et al., 2011; Margulis, 2013; Margulis & Simchy-Gross, 2016; Rowland et al., 2019), then the repetition account could explain the music-like percepts observed at all list lengths in Study 1, the music-like percepts reported by Kaminska and Mayer (2002) in their study of the verbal transformation effect, the song-like percepts for speech-like and noise-like stimuli in Study 2, the sound to music percepts in Studies 4 and 5, and the similar time course of the speech to song illusion and the sound to music illusion in Studies 3-5.

It is not clear, however, how the repetition account could explain the lexical transformations and patternization percepts reported in Study 1, the lexical transformations reported by Kaminska and Mayer (2002) in their study of the verbal transformation effect, the influence of phonological neighborhood density on the speech to song illusion in Study 2 (see also Castro et al. 2018 and Vitevitch et al., 2021), or why emotional sounds can induce the sound to music illusion as in Study 5, but emotional words do not affect the speech to song illusion as in Vitevitch et al. (2021).

In the Node Structure Theory account, presentation of a word or phrase initially primes and activates lexical nodes associated with those words, bringing to conscious awareness the information associated with the word and a speech-like percept. With repeated activation of the same lexical nodes, satiation occurs resulting in the lexical nodes being temporarily unable to accumulate priming and be activated, and in the inability to retrieve information associated with that word/lexical node. In the case of the speech to song illusion, additional presentations of the stimulus continue to prime the syllable nodes, making salient the rhythmic pattern in the repeated phrase, and resulting in a song-like percept (Castro et al., 2018; Mullin, et al., 2021; Vitevitch et al., 2021). In the case of the verbal transformation effect, the lexical node associated with the repeated word satiates, giving another lexical node that is similar to the input the opportunity to be primed and ultimately activated by the repeated stimulus, bringing to conscious awareness another word (MacKay, et al., 1993; Shoaf & Pitt, 2002).

Thus, the NST account can easily explain the music-like percepts observed at all list lengths in Study1, the music-like percepts reported by Kaminska and Mayer (2002) in their study of the verbal transformation effect, the lexical transformations and patternization percepts reported in Study 1, the lexical transformations reported by Kaminska and Mayer (2002) in their study of the verbal transformation effect, the influence of phonological neighborhood density on the speech to song illusion in Study 2 (see also Castro et al. 2018 and Vitevitch et al., 2021), and why the emotional qualities of words used as stimuli do not affect the speech to song illusion as in Vitevitch et al. (2021). What NST cannot easily explain is the song-like percepts for the noise-like stimuli in Study 2, the sound to music percepts in Studies 4 and 5, or the similar time course of the speech to song illusion and the sound to music illusion in Studies 3-5. Although environmental sounds do influence speech production (Mädebach, Wöhner, Kieseler &

Jescheniak, 2017), and word recognition (Toon & Kukona, 2020) it is not clear in the context of NST how environmental sounds would transform to music-like percepts.

We acknowledge that the present set of studies have raised more questions than they have answered. For example, the results of Study 1 have blurred the distinction between the verbal transformation effect and the speech to song illusion, leading us to question whether they are actually distinct illusions, or whether researchers only view them as distinct due to the methods, tasks, stimuli, and participant response options typically used to examine one or the other illusion. We also question whether the tendency of researchers to focus only on a specific phenomenon in a given study (e.g., studying VTE, but not the speech to song illusion) also contributed to the (mis)conception that these are distinct illusions. Clearly, additional studies—using different methods, a broader range of methods, and combinations of methods—will be required to determine more definitively if the auditory illusions we examined here are indeed distinct illusions.

If the auditory illusions we examined here are indeed distinct illusions, then what is the "boundary" that separates one illusion from another? The boundary between illusions could be related to the amount of time that listeners are exposed to the stimulus (compare the 10 repetitions typically used to examine the speech to song illusion to the 4 minutes typically used to examine the verbal transformation effect). Perhaps certain percepts only arise when exposed to a stimulus for a short amount of time, whereas other percepts will only emerge when exposed to a stimulus for a longer amount of time. The boundary between illusions could also be related to some characteristic of the stimulus. Perhaps certain percepts only arise for single, monosyllabic words compared to phrases containing several multisyllabic words (see Study 1 & Castro et al.,

2018). Alternatively, perhaps the boundary lies between speech sounds and non-speech sounds, or between a perceptual/"lower-level" mechanism and a cognitive/"higher-level" mechanism.

A possible solution: Two mechanisms

Neither the repetition account nor the NST account adequately explains all of the phenomena associated with the auditory illusions examined here. Does that suggest that neither the repetition account nor the NST account is the *right* single mechanism to explain the auditory illusions we examined here? In their study of the verbal transformation effect Kaminska and Mayer (2002) noted that the speech-specific account of the VTE (Warren, 1983) could not account for the transformations they observed for nonspeech sounds. Instead, they proposed a multi-dimensional network with different types of representations (i.e., verbal and nonverbal information) that has spreading-activation and a perceptual criterion that can shift under different listening conditions. When speech is heard under natural conditions the verbal representations drive perception and cognition. However, under the "ecologically invalid listening conditions of the transformation paradigm" (Kaminska & Mayer, 2002; pg. 328) the criterion is lowered, allowing other indirectly activated representational units (i.e., nonverbal information) to emerge as percepts.

Note that the first report of the speech to song illusion (Deutsch et al., 2011) and the first report of the sound to music illusion (Simchy-Gross & Margulis, 2018) were published several years after this model was proposed by Kaminska and Mayer (2002), so Kaminska and Mayer could not have accounted for these phenomena when formulating their model. As such, it is not clear if the model proposed by Kaminska and Mayer (2002) can also account for the speech to song illusion and the sound to music illusion. One point to consider in assessing whether the model proposed by Kaminska and Mayer (2002) can account for the speech to song illusion and

the sound to music illusion is whether the 10 repetitions of a phrase in these illusions is equal in ecological invalidity to the 4 minutes of stimulus repetition used in the VTE paradigm to shift the perceptual criterion in the multi-dimensional network proposed by Kaminska and Mayer (2002).

Although we are proponents of various types of network models (e.g., Vitevitch, 2022; Vitevitch & Storkel, 2013), we take a different approach in the present case. Instead, we consider the possibility that the present findings may suggest that a *single* mechanism cannot explain all of the auditory illusions we examined here. Just as two mechanisms are required to fully explain all aspects of color perception in vision, in what follows we describe how two mechanisms might be needed to fully explain all aspects of the auditory illusions we examined here.

Given the different evolutionary timelines for music and language processing, and the evidence for distinct cortical pathways for processing music and language we propose two mechanisms to fully account for the auditory illusions examined here. One mechanism is perceptual/"lower-level" in nature, and the second mechanism is cognitive/"higher-level" in nature. The perceptual/"lower-level" mechanism is music-based, whereas the cognitive/"higher-level" is language-based.

We describe the music-based mechanism as being perceptual/"lower-level" in nature in part because very-low frequency sounds that are below or near auditory thresholds and not consciously detected have been shown to induce increased head movements associated with dancing at a live concert (Cameron et al., 2022). Inducing rhythmic movements with sounds that are near or below the perceptual threshold and not consciously detected suggests that the music-based mechanism involved in the auditory illusions examined here could be perceptual/"lower-level" in nature.

We also describe the music-based mechanism as being perceptual/"lower-level" in nature because a very basic acoustic parameter—amplitude modulation (AM)—has been shown to be important for distinguishing music from speech (Chang, Teng, Assaneo & Poeppel, 2022). Chang et al. (2022) further found that more musically sophisticated participants were more likely to judge the sounds in a music detection task as being music despite that fact that 50% of the AM stimuli were music, and the remaining 50% were some other sound. Furthermore, musical experience did not significantly influence performance in an analogous speech detection task where 50% of the AM stimuli were speech, and the remaining 50% were some other sound. Chang et al. noted that musical aptitude has also been shown to influence ratings in the speech to song illusion (Rowland et al., 2018; Tierney et al., 2018; Tierney et al., 2021; Vanden Bosch der Nederlanden et al., 2015). They further suggested that the repetition of phrases as in the speech to song illusion could produce an amplitude modulated spectrum that could contribute to the increased song-likeness ratings in speech to song illusion tasks.

We propose that a perceptual/"lower-level" music-based mechanism—such as the AM detection system described by Chang et al (2022)—could be the mechanism that contributes to the emergence of music-like percepts in the speech to song illusion, in the verbal transformation effect (as in Study 1 and in Kaminska & Mayer, 2002), with verbal stimuli like those used in Study 1 that are "between" the stimuli typically used in the verbal transformation effect and in the speech to song illusion, and when non-speech, environmental sounds are repeated (i.e., the nonspeech sounds used by Kaminska & Mayer, 2002, and in the sound to music illusion).

We are not alone in proposing that a perceptual/"lower-level" music-based mechanism may lead to music-like percepts in certain auditory illusions. Indeed, Rowland et al. (2019; pg. 588) suggest:

> *The apparent ubiquity of repetition-induced perceived musical attributes using different acoustic and environmental categories suggests a general mechanism not specifically tied to speech, or any particular component (spectral or rhythmic) in the signal. The robust illusion described by Deutsch et al. (2011) may be a special case of a broader phenomenon encompassing generalized repeated auditory material, better described as a "repetition-to-music" effect.*

Where we differ from Rowland et al. (2019) and others is in suggesting that a second mechanism—one that is language-based and cognitive/"higher-level" in nature—is *also* involved in producing the various percepts reported in the auditory illusions examined here: verbal transformation effect, speech to song illusion, and sound to music illusion.

We propose that the cognitive/"higher-level" language-based mechanism that is also involved in the verbal transformation effect, speech to song illusion, and sound to music illusion is a mechanism similar to the mechanisms proposed in Node Structure Theory (MacKay, 1987). Recall that NST has previously been put forward as an account of the verbal transformation effect (MacKay et al., 1993; Shoaf & Pitt, 2002) and of the speech to song illusion (Castro et al., 2018; Mullin, et al., 2021; Vitevitch et al., 2021). The satiation of lexical nodes described in NST not only accounts for lexical transformations (and may also account for the new percept, patternization, that we discovered in Study 1), but may also allow music-like percepts to emerge by allowing attention to shift to the musical qualities associated with syllable nodes or to the lower-level AM system. Musical experience may influence whether attention stays at the syllable nodes or shifts further to the lower-level AM system, thus accounting for the influences that musical training has on song-likeness ratings in the speech to song illusion. The extent to which the AM system is engaged or attended to may also account for why some stimuli better evoke the speech to song illusion than others.

Previous descriptions of the NST account of the speech to song illusion fell short in two critical areas: (1) explaining why music-like percepts emerged from repetition of phrases/word

lists instead of lexical transformations, and (2) explaining how the repetition of non-speech sounds also produced music-like percepts. As observed in Study 1, musical percepts emerged for shorter word lists, not just longer word lists, suggesting that the verbal transformation effect and the speech to song illusion may not be as distinct as previous studies might have suggested. More importantly, it was observed in Study 1 that (when given the chance) participants do report lexical transformations for longer word-lists (i.e., phrases) like those typically used to examine the speech to song illusion. This again suggests that these two auditory illusions may not be as distinct as previous studies might have suggested. Note that NST still cannot explain how the repetition of non-speech sounds produces music-like percepts.

However, by themselves, music-based accounts of these auditory illusions are not able to explain how lexical transformations emerge when spoken stimuli are repeated. Further, music-based accounts of the speech to song illusion are not able to explain how phonological variables such as neighborhood density (Study 2 and Castro et al., 2018) and phonological clustering coefficient (Vitevitch et al., 2021) influence the speech to song illusion. Furthermore, music-based accounts of the speech to song illusion and the sound to music illusion cannot explain why "emotion" does not affect the speech to song illusion (Vitevitch et al., 2021), but does elicit the sound to music illusion (as in Study 5).

Therefore, no *single* mechanism (at least no single mechanism discussed in this article) can account for all of the phenomena related to the auditory illusions examined here. This may mean that the right single mechanism has not been discovered yet. Alternatively, it may mean that two mechanisms—a perceptual/"lower-level" music-based mechanism and a cognitive/"higher-level" language-based mechanism—may both be required to completely account for all of the phenomena related to the auditory illusions examined here. We recognize

that the theoretical parsimony of a single mechanism is lost by our proposal for two mechanisms. However, there is also much to be gained with our current proposal.

For example, having two mechanisms involved in processing may be the only way to explain why some illusions give you both music-like percepts and lexical transformations, such as the verbal transformation effect and speech to song illusion as observed in Study 1. Similarly, having two mechanisms involved in processing may explain why experience with music influences some processes, such as AM discrimination of speech/music (Chang et al., 2022) and some auditory illusions (e.g., speech to song illusion), but not others. In short, proposing two mechanisms involved in processing may actually be more parsimonious than continuing to add caveats and exceptions to a single mechanism to account for all of the phenomena related to these auditory illusions.

Another thing that can be gained by considering that two mechanisms are involved in the auditory illusions examined here is that these proposed mechanisms are "pre-existing" mechanisms that are involved in other cognitive processes, namely music and language process. Instead of developing *ad hoc* accounts of each auditory illusion, the proposal we advance for two mechanisms being involved in the auditory illusions that we examined ties all of the illusions to a richer and broader theoretical literature. By eschewing *ad hoc* accounts of each auditory illusion and instead connecting the auditory illusions to music- and language-based processes that already exist we gain the opportunity to learn more about how those perceptual and cognitive systems work under typical conditions.

As an example of how new insights can be gained by connecting to the research literature in auditory perception and language processing in general, consider the sound to music illusion (Simchy-Gross & Margulis, 2018) and the music-like percepts reported in the study of the verbal

transformation effect by Kaminska and Mayer (2002). In both illusions, environmental sounds are repeated and take on a music-like quality. Given the blurred distinction between the verbal transformation effect and the speech to song illusion that was observed in Study 1, one might wonder if the "reverse" is possible. That is, can non-speech sounds such as "bad electronic music," or "radio interference" also be perceived as speech (https://haskinslabs.org/research/features-and-demos/sinewave-synthesis)? Indeed, there is a rich literature on sine wave speech (Remez, Rubin, Pisoni & Carrell, 1981), in which words, phrases, and sentences are perceived when participants are presented with three or four time-varying sinusoids. Thus, connecting to the broader literature in auditory perception and language processing in general could lead to increased understanding of the systems involved as well as new predictions about the auditory illusions.

One example of how connecting to the broader literature in auditory perception and language processing in general could lead to new predictions about these auditory illusions is derived from previous findings that show that musical experience influences song-like ratings. Given that "experience" with music influences song-like ratings, might "experience" with language influence the lexical transformations experienced in the verbal transformation effect (and perhaps in the speech to song illusion as in Study 1)? For example, do people with larger vocabularies experience more lexical transformations than people with smaller vocabularies? Similarly, given the phonological overlap of words in various languages (Vitevitch, 2012), do speakers of more than one language experience lexical transformations of a word in one language to a word in another language that they know when experiencing the verbal transformation effect?

We believe that music perception and acoustic processing also stand to gain from anchoring the auditory illusions we examined to the music- and language-based processes that we proposed. We eagerly await future studies to test the hypotheses we have put forward, and to test new hypotheses derived from the music- and language-based processes that might be responsible for producing these auditory illusions.

Acknowledgements

We thank Sarah K. Brummett and Maddie Kentch for their assistance in collecting data in several of the studies reported here.

Disclosure Statement

The authors report there are no competing interests to declare.

Data availability statement

The data associated with this paper are available upon request from the corresponding author.

References

Arjmand, H-A., Hohagen, J., Paton, B. & Rickard, N.S. (2017). Emotional responses to music: Shifts in frontal brain asymmetry mark periods of musical change. *Frontiers in Psychology, 8.* https://www.frontiersin.org/articles/10.3389/fpsyg.2017.02044

Arnal, L. H., Flinker, A., Kleinschmidt, A., Giraud, A.-L., & Poeppel, D. (2015). Human screams occupy a privileged niche in the communication soundscape. *Current Biology, 25*, 1-6.

Boersma, P. & Weenink, D (1992). Praat: doing phonetics by computer [Computer program]. Version 6.1.15, retrieved 15 April 2020 from https://www.praat.org.

Cameron, D.J., Dotov, D., Flaten, E., Bosnyak, D., Hove, M.J. & Trainor, L.J. (2022). Undetectable very-low frequency sound increases dancing at a live concert. *Current Biology, 32*, R1222-R1223.

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): e0198656. https://doi.org/10.1371/ journal.pone.0198656

Chang, A., Teng, X., Assaneo, M.F. & Poeppel, D. (2022, November 6). Amplitude modulation perceptually distinguishes music and speech. https://doi.org/10.31234/osf.io/juzrh

Cohen J. D., MacWhinney B., Flatt M., & Provost J. (1993). PsyScope: A new graphic interactive environment for designing psychology experiments. *Behavioral Research Methods, Instruments, and Computers, 25*, 257–271.

Cutler A. (1991) Linguistic rhythm and speech segmentation. In: Sundberg J., Nord L., Carlson R. (eds) *Music, Language, Speech and Brain*. Wenner-Gren Center International Symposium Series. Palgrave, London.

Deutsch, D. (1995). *Musical Illusions and Paradoxes*. La Jolla: Philomel Records.

Deutsch, D., Henthorn, T. & Lapidis, R. (2011). Illusory transformation from speech to song. *Journal of the Acoustical Society of America, 129*, 2245-2252.

Dorman, M.F., Loizou, P.C., Fitzke, J. & Tu, Z. (1998). The recognition of sentences in noise by normal-hearing listeners using simulations of cochlear-implant signal processors with 6–20 channels. *The Journal of the Acoustical Society of America, 104* (6), 3583-3585. https://doi.org/10.1121/1.423940

Dorman, M.F., Natale, S.C., Baxter, L., et al. (2020). Approximations to the voice of a cochlear implant: Explorations with single-sided deaf listeners. *Trends in Hearing*, 24 doi:10.1177/2331216520920079

Dziubalska-Kołaczyk, K. (2002). *Beats-and-Binding Phonology*. Frankfurt am Main: Peter Lang.

Eleuteri, V., et al. (2022). The form and function of chimpanzee buttress drumming, *Animal Behaviour*, https://doi.org/10.1016/j.anbehav.2022.07.013

Falk, S., Rathcke, T., & Dalla Bella, S. (2014). When speech sounds like music. *Journal of Experimental Psychology: Human Perception and Performance, 40*(4), 1491–1506. https://doi.org/10.1037/a0036858

Franconeri, S. L., & Simons, D. J. (2003). Moving and looming stimuli capture attention. *Perception & Psychophysics, 65*, 999–1010.

Fujii, S. & Wan, C.Y. (2014). The Role of Rhythm in Speech and Language Rehabilitation: The SEP Hypothesis. *Frontiers in Human Neuroscience, 8*, 777.

Haiduk, F. & Fitch, W.T. (2022). Understanding Design Features of Music and Language: The Choric/Dialogic Distinction. *Frontiers in Psychology, 13*, 786899.

Harrison, L. & Loui., P. (2014). Thrills, chills, frissons, and skin orgasms: toward an integrative model of transcendent psychophysiological experiences in music. *Frontiers in Psychology, 5*:790. doi:10.3389/fpsyg.2014.00790

Hauser, M.D., Yang, C., Berwick, R.C., Tattersall, I., Ryan, M. J., Watumull, J., Chomsky, N. & Lewontin, R.C. (2014). The mystery of language evolution. *Frontiers in Psychology, 5,* 00401.

Hering, E. (1872). "Zur Lehre vom Lichtsinne". *Sitzungsberichte der Mathematisch–Naturwissenschaftliche Classe der Kaiserlichen Akademie der Wissenschaften*. K. K. Hofund Staatsdruckerei in Commission bei C. Gerold's Sohn. LXVI. Band (III Abtheilung).

Honda, S., Ishikawa, Y., Konno, R., Imai, E., Nomiyama, N., Sakurada, K., Koumura, T., Kondo, H.M., Furukawa, S., Fujii, S. & Nakatani, M. (2020). Proximal Binaural Sound Can Induce Subjective Frisson. *Frontiers in Psychology, 11*:316. doi: 10.3389/fpsyg.2020.00316

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

Jaisin, K., Suphanchaimat, R., Figueroa Candia, M.A., and Warren, J.D. (2016). The speech-to-song illusion is reduced in speakers of tonal (vs. non-tonal) languages. Frontiers in Psychology, 7. https://www.frontiersin.org/articles/10.3389/fpsyg.2016.00662

JASP Team (2022). JASP (Version 0.16.3) [Computer software].

Kaminska Z., & Mayer P. (2002). Changing words and changing sounds: A change of tune for verbal transformation theory? *European Journal of Cognitive Psychology, 14*, 315–333.

Kennedy, R., Clifford, S., Burleigh, T., Waggoner, P., Jewell, R., & Winter, N. (2020). The shape of and solutions to the MTurk quality crisis. *Political Science Research and Methods*, 8(4), 614-629.

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

MacKay D. G., Wulf G., Yin C., & Abrams L. (1993). Relations between word perception and production: New theory and data on the verbal transformation effect. *Journal of Memory and Language, 32,* 624– 646.

Mädebach, A., Wöhner, S., Kieseler, M.-L., & Jescheniak, J. D. (2017). Neighing, barking, and drumming horses—object related sounds help and hinder picture naming. *Journal of Experimental Psychology: Human Perception and Performance, 43(9)*, 1629–1646.

Margulis, E. H. (2013). Repetition and emotive communication in music versus speech. *Frontiers in Psychology, 4*, 167.

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

McGuire, A.B., Gillath, O. & Vitevitch, M.S. (2016). Effects of mental resource availability on looming task performance. *Attention, Perception & Psychophysics, 78*, 107-113.

Mullin, H.A.C., 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*, 1281–1296.

Ramus, F., Nespor, M. & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition, 73*, 265-292.

Remez, R.E., Rubin, P.E., Pisoni, D.B., & Carrell, T.D. (1981). Speech perception without traditional speech cues. *Science, 212*, 947-950.

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*, 583–590.

Siew, C.S.Q. & Vitevitch, M.S. (2016). Spoken word recognition and serial recall of words from components in the phonological network. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42*, 394-410.

Simchy-Gross, R. & Margulis, E.H. (2018). The sound-to-music illusion: Repetition can musicalize nonspeech sounds. *Music & Science, 1*, 1-6.

Shoaf, L.C. & Pitt, M.A. (2002). Does node stability underlie the verbal transformation effect? A test of node structure theory. *Perception & Psychophysics, 64*(5):795-803.

Soehlke, L.E., Kamat, A., Castro, N. & Vitevitch, M.S. (2022). The influence of memory on the speech-to-song illusion. *Memory & Cognition*. https://doi.org/10.3758/s13421-021-01269-9

Thompson, R. F., & Spencer, W. A. (1966). Habituation: a model phenomenon for the study of neuronal substrates of behavior. *Psychological Review, 73*, 16-43.

Tierney, A., Patel, A. D., & Breen, M. (2018). Acoustic foundations of the speech-to-song illusion. *Journal of Experimental Psychology: General, 147*(6), 888–904.

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), 1681-1697.

Toon, J. & Kukona, A. (2020). Activating Semantic Knowledge During Spoken Words and Environmental Sounds: Evidence from the Visual World Paradigm. *Cognitive Science, 44*: e12810.

Vitevitch, M.S. (2012). What do foreign neighbors say about the mental lexicon? *Bilingualism: Language & Cognition, 15*, 167-172.

Vitevitch, M.S. (2022). What Can Network Science Tell Us About Phonology and Language Processing? *Topics in Cognitive Science, 14*, 127-142.

Vitevitch, M.S. and Aljasser, F.M. (2021). Phonotactics in Spoken-Word Recognition. In *The Handbook of Speech Perception* (eds J.S. Pardo, L.C. Nygaard, R.E. Remez and D.B. Pisoni). https://doi.org/10.1002/9781119184096.ch11

Vitevitch, M.S. & Luce, P. (2016). Phonological neighborhood effects in spoken word perception and production. *Annual Review of Linguistics, 2*, 75-94.

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*, 585-597.

Vitevitch, M.S., Siew, C.S.Q., Castro, N., Goldstein, R., Gharst, J.A., Kumar, J.J., and Boos, E.B. (2015). Speech error and tip of the tongue diary for mobile devices. *Frontiers in Psychology, 6*:1190. doi: 10.3389/fpsyg.2015.01190

Vitevitch M. S., Stamer M. K., & Sereno J. A. (2008). Word length and lexical competition: Longer is the same as shorter. *Language and Speech, 51*, 361–383.

Vitevitch, M.S. & Storkel, H.L. (2013). Examining the acquisition of phonological word forms with computational experiments. *Language & Speech, 56*, 491-527.

Vitevitch, M.S., Storkel, H.L., Francisco, A.C. Evans, K.J. & Goldstein, R. (2014). The influence of known-word frequency on the acquisition of new neighbors in adults: evidence for exemplar representations in word learning. *Language, Cognition and Neuroscience, 29*, 1311-1316.

Warren, R. M. (1983). Auditory illusions and their relation to mechanisms normally enhancing accuracy of perception. *Journal of the Audio Engineering Society, 31*(9), 623-629.

Warren, R. M., & Gregory, R. L. (1958). An auditory analogue of the visual reversible figure. *The American Journal of Psychology, 71,* 612-613.

Winkler, A., Kogan V.V. & Reiterer, S.M. (2023). Phonaesthetics and personality-Why we do not only prefer Romance languages. *Frontiers in Language Sciences, 2*, https://www.frontiersin.org/articles/10.3389/flang.2023.1043619

Xu, J., Guo, X., Liu, M., Xu, H. & Huang, J. (2023). Self-construal priming modulates sonic seasoning. *Frontiers in Psychology, 14*, https://www.frontiersin.org/articles/10.3389/fpsyg.2023.1041202

Young, T. (1802). The Bakerian Lecture. On the theory of light and colours. *Philosophical Transactions of the Royal Society of London*, *92*, 12-48.

Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology, 9*, 1–27.

Zhang, S. (2011, August). *Speech-to-song illusion in MC: Acoustic parameter vs. perception.* Poster presented at the biennial meeting of the Society for Music Perception and Cognition, Rochester, NY.