HHS Public Access

Author manuscript

Q J Exp Psychol (Hove). Author manuscript; available in PMC 2021 July 19.

Published in final edited form as:

Q J Exp Psychol (Hove). 2021 April ; 74(4): 585–597. doi:10.1177/1747021820969144.

Phonological but not semantic influences on the speech-to-song illusion

Michael S Vitevitch[1], **Joshua W Ng**[1], **Evan Hatley**[1], **Nichol Castro**[2]

[1]University of Kansas, Lawrence, KS, USA

[2]University at Buffalo, Buffalo, NY, USA

Abstract

In the speech to song illusion, a spoken phrase begins to sound as if it is being sung after several repetitions. Castro et al. (2018) used Node Structure Theory (NST; MacKay, 1987), a model of speech perception and production, to explain how the illusion occurs. Two experiments further test the mechanisms found in NST—priming, activation, and satiation—as an account of the speech to song illusion. In Experiment 1, words varying in the phonological clustering coefficient influenced how quickly a lexical node could recover from satiation, thereby influencing the song-like ratings to lists of words that were high versus low in phonological clustering coefficient. In Experiment 2, we used equivalence testing (i.e., the TOST procedure) to demonstrate that once lexical nodes are satiated the higher level semantic information associated with the word cannot differentially influence song-like ratings to lists of words varying in emotional arousal. The results of these two experiments further support the NST account of the speech to song illusion.

Keywords

Speech to song; illusion; node structure theory; language

Introduction

Perceptual illusions occur when our percept does not match what is actually in the environment. As “the dress” (https://en.wikipedia.org/wiki/The_dress) and the Yanny-Laurel debate (https://en.wikipedia.org/wiki/Yanny_or_Laurel) of recent internet fame suggest, visual and auditory illusions capture the interest of the general public. In addition to entertaining the general public and zoo animals (Regaiolli et al., 2019), perceptual illusions provide researchers with another way to examine the limits of the perceptual and cognitive systems involved in various illusions, thereby increasing our fundamental understanding of

Corresponding author: Michael S Vitevitch, Department of Psychology, University of Kansas, Lawrence, KS 66045-7556, USA., mvitevit@ku.edu.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Supplementary material

The supplementary material is available at: qjep.sagepub.com

these systems (Gregory, 1968; see also Boyette et al., 2020; Vitevitch, 2003; Vitevitch & Donoso, 2011; Vitevitch & Siew, 2017).

The speech to song (*S2S*) illusion is an auditory illusion that occurs when a spoken phrase is repeated several times resulting in the phrase sounding like it is being sung instead of spoken (Deutsch et al., 2011). The illusion occurs not only with English phrases for native speakers of English but also with phrases in Mandarin for native speakers of Mandarin (Zhang, 2011) and with phrases in German for native speakers of German (Falk & Rathcke, 2010). Functional magnetic resonance imaging revealed that brain regions associated with pitch processing and song production tend to be activated when listeners experience the illusion (Tierney et al., 2013), further attesting to the robustness of this illusion (see also Vanden Bosch der Nederlanden et al., 2015).

One account for how the S2S illusion occurs points to the important role that repetition of the stimulus plays in eliciting the illusion (Margulis, 2013; Margulis & Simchy-Gross, 2016; Rowland et al., 2019). However, accounts that appeal only to repetition as the mechanism that is responsible for the illusion fail to explain why the initial percept is that of speech, what exactly repetition does to change the percept, and why the subsequent percept is music-like instead of something else, such as nonsensical babble as occurs when speakers repeatedly produce a word and experience semantic satiation (Lambert & Jakobovits, 1960). Furthermore, the accounts that appeal only to repetition appear *ad hoc* in that they fail to connect the illusion to other auditory illusions or perceptual phenomena and do not situate the S2S illusion into a richer theoretical account of perception or cognition more broadly.

Clearly, repetition of the stimulus to the listener is *necessary* for the S2S illusion to occur. However, repetition alone is not *sufficient* to explain how or why the S2S illusion occurs.

In contrast to the repetition account of the S2S illusion, recent work by Castro et al. (2018) suggests that the mechanisms found in Node Structure Theory (*NST*; MacKay, 1987), a model of language perception and production, may be able to explain the illusion in such a way that connects it to other auditory illusions and to other perceptual and cognitive phenomena, thereby providing insight into the inner workings of these perceptual and cognitive systems. Indeed, NST has been used to account for word retrieval and production (MacKay, 1987), tip-of-the-tongue states (Burke et al., 1991), differences in language processing due to aging (e.g., MacKay & Burke, 1990), the language production deficits of H.M. (MacKay et al., 1998), and another auditory illusion known as the Verbal Transformation Effect (*VTE*; MacKay et al., 1993).

NST uses nodes to represent phonemes, syllables, words, and other types of linguistic information. Links connect constituent nodes together such that phoneme nodes connect to syllable nodes, and syllable nodes connect to lexical nodes. The nodes are organised into different systems—including the sentential system and the phonological system—with nodes linked within and across systems. For example (see Figure 1), the node for the word *frisbee* connects at the higher level to the semantic node “frisbees are thrown” and connects at lower levels to nodes for the syllables *fris* and *bee*, which then connect to the phonological nodes /f/, and so on. The phonological nodes would eventually connect to nodes representing

motor-related information to articulate a word or phrase (*N.B.*, several nodes at higher and lower levels are not included in Figure 1 to simplify the illustration).

In NST, language production and perception are made possible by three fundamental mechanisms: *priming, activation, and satiation. Priming* (similar to spreading activation in other models) is transmitted along links to increase activity in a node. During speech perception, incoming acoustic–phonetic information primes various phonological nodes, based on the extent to which the nodes match the input. When a node accumulates enough priming (summed across its connections and over time) to surpass an activation threshold, it is said to be *activated* in NST. Activation of a node is "all-or-none" and results in conscious awareness of the information represented by that node.

An important exception to the "all-or-none" principle of activation in NST occurs during speech perception where it is typically sufficient to prime, but not fully activate units in the phonological system (including syllable nodes and nodes representing individual phonemes). The priming, but "failure" to fully activate syllable and phoneme nodes results in the listener perceiving words (which are primed and activated, and therefore in conscious awareness) rather than sequences of phonemes when listening to speech (MacKay, 1987).

Satiation refers to the temporary reduction in the ability of a node to accumulate priming and be activated. This state is induced by repeated activation of the same node and is believed to serve the evolutionary function of bringing new stimuli to awareness instead of keeping old or unchanging information in awareness (MacKay, 1987).

Only nodes that have been activated experience satiation. Therefore, lexical nodes that are activated can become satiated if stimulated repeatedly, but syllable and phoneme nodes which are primed but not activated during speech perception would not experience satiation. When a given node is satiated, the "most-primed-wins" principle results in another related and highly primed node being activated instead (MacKay, 1987; MacKay et al., 1993).

The six experiments reported by Castro et al. (2018) examined how the mechanisms of priming, activation, and satiation (as well as the "most-primed-wins" principle) explain the S2S illusion. In the S2S illusion, the spoken phrase first primes and activates lexical nodes, giving the initial percept of speech. Repetition of the phrase causes the lexical nodes associated with the words in that phrase to satiate (i.e., they can no longer be activated), resulting in the loss of the initial speech percept.

Importantly, however, despite satiation of the lexical nodes, the syllable nodes continue to be primed by the repeating phrase. Recall that priming without activation in the syllable nodes occurs during everyday language perception, where one perceives words (because lexical nodes are activated) instead of a sequence of syllables or phonemes (because the syllable and phoneme nodes are only primed and not activated). It is widely recognised that syllables are a unit of rhythmical structure in speech (e.g., Cutler, 1991; Fujii & Wan, 2014; Jackendoff, 2009; Ramus et al., 1999). With satiation of the lexical nodes, the speech percept is lost. However, the continued priming of the syllable nodes and the "most-primed-wins" principle makes salient the metrical pattern found in the repeated phrase, producing the song-like quality experienced by listeners in the S2S illusion.

We report in what follows the results of two experiments that further examine how the mechanisms of priming, activation, and satiation found in NST explain the S2S illusion. In Experiment 1, we manipulated a phonological characteristic of words to differentially affect the priming and activation mechanisms of NST (although satiation is obviously involved in producing the illusion). In Experiment 2, we manipulated a semantic characteristic of words to demonstrate more clearly the important role that satiation of the lexical nodes plays in the S2S illusion.

Experiment 1

In Experiment 1 of Castro et al. (2018), the S2S illusion was elicited using lists of four words that were recorded in isolation, then concatenated together and repeated for listeners. Using lists of words demonstrated that it is not necessary to extract a phrase from a sentence as had been previously done to elicit the S2S illusion.

Using lists of words also enabled Castro et al. to manipulate various characteristics about the words to examine how those variables might influence the S2S illusion. Specifically, the words in the lists they used varied in *phonological neighbourhood density*, which measures the number of words that sound like a given word (Vitevitch & Luce, 2016). A word that has many words that sound similar to it is said to have a dense phonological neighbourhood, whereas a word that has few words that sound similar to it is said to have a sparse phonological neighbourhood. Phonological neighbourhood density has been shown to influence certain aspects of speech perception, spoken word recognition, speech production, word-learning, and various aspects of memory (for a review see Vitevitch & Luce, 2016), as well as the tip-of-the-tongue state (Vitevitch & Sommers, 2003) and the VTE (Bashford et al., 2006)—two phenomena that NST also accounts for.

Castro et al. found that lists containing words with dense phonological neighbourhoods elicited higher song-like ratings than lists containing words with sparse phonological neighbourhoods after repetition. They suggested that the priming transmitted by phonological nodes to the lexical nodes of words with dense phonological neighbourhoods would be distributed among more words than the priming transmitted by phonological nodes to the lexical nodes of words with sparse phonological neighbourhoods. The different amounts of priming being transmitted to words with dense neighbourhoods versus words with sparse neighbourhoods meant that words in sparse phonological neighbourhoods would recover from satiation and be activated more quickly than lexical nodes for words with dense phonological neighbourhoods. Activation of the lexical nodes for words with sparse phonological neighbourhoods would bring the speech percept back to awareness, and decrease the song-like percept (and ratings) for such words. Lexical nodes for words with dense phonological neighbourhoods would remain satiated, with their associated syllable nodes continuing to be primed and maintaining a music-like percept (and therefore result in a higher song-like rating).

To further test the mechanisms of priming, activation, and satiation (and the “most-primed wins” principle) found in NST as a way to account for the S2S illusion, we examined in the present experiment how a measure of phonological similarity derived from network science

—namely, the *clustering coefficient* (Chan & Vitevitch, 2009)—affects the ability of lexical nodes to recover from satiation. In a network of phonologically related words (Vitevitch, 2008), the clustering coefficient, *C*, measures the extent to which neighbours of a word are also neighbours of each other. As illustrated in Figure 2, the words *badge* and *log* have the same number of phonological neighbours (the variable manipulated in Castro et al., 2018), but the phonological neighbours of *badge* are also neighbours with each other to a greater extent than the phonological neighbours of *log*.

In numerous psycholinguistic experiments, the clustering coefficient has been shown to influence speech perception (Chan & Vitevitch, 2009), speech production (Chan & Vitevitch, 2010), short-term memory (Vitevitch et al., 2012), and word learning (Goldstein & Vitevitch, 2014). Chan and Vitevitch (2009) suggested that the influence of *C* on lexical processes could be modelled as activation that diffused across a network composed of nodes representing words in the lexicon and connections between phonologically related words. The verbal model proposed by Chan and Vitevitch (2009) was subsequently confirmed via computer simulations (Siew, 2019; Vitevitch et al., 2011). To facilitate describing the predictions of how *C* might influence the S2S illusion, we will henceforth use the terminology employed in NST (i.e., "priming being transmitted") instead of "activation spreading" across a network-like structure.

We expect that high *C* words (*badge* in Figure 2) will be primed, activated, and then satiated just like low *C* words (*log* in Figure 2). Similar to the manipulation of phonological neighbourhood density in Experiment 1 in Castro et al. (2018), however, we expect there will be a difference in how quickly words with high versus low *C* will recover from satiation due to how quickly priming can accumulate to again surpass the activation threshold.

For low *C* words (*log* in Figure 2), the small number of interconnections among the neighbours will result in some priming from the neighbours being transmitted back to the target word, but most of the priming will be transmitted (or dispersed) to the rest of the network (i.e., words related to the neighbours of *log*, but not shown in Figure 2). With little priming accumulating over time and across connections words with low *C*—once activated and satiated—will stay satiated longer than high *C* words. With lexical nodes corresponding to the low *C* words staying satiated longer, the song-like percept emerging from priming of the syllable nodes (which carry the rhythmic information of language) will persist and result in higher song-like ratings for lists of words with low *C* compared to lists of words with high *C*.

In the case of high *C* words, where the neighbours are highly interconnected with each other, priming will be "trapped" among the interconnected neighbours and transmitted back to the target word rather than be dispersed to the rest of the network as happens for low *C* words (as observed in the simulations by Vitevitch et al., 2011). With the trapped priming being summed over time from the highly interconnected neighbours, a word with high *C* could recover from satiation and be activated again more quickly than a word with low *C*. The activation of lexical nodes corresponding to the high *C* words will again bring to conscious awareness the speech percept, thereby weakening the song-like percept and resulting in

lower song-like ratings for lists containing words with high *C* compared to lists containing words with low *C*.

To test these predictions, we used the same task used in Castro et al. (2018) and in many other studies of the S2S illusion. Participants listened to 10 repetitions of each list of words. At the end of the repetitions, the participants provided a rating on a 5-point Likert-type scale with 1 corresponding to "sounds like speech" and 5 corresponding to "sounds like song." Higher ratings on the scale indicate experiencing a song-like percept, whereas lower ratings on the scale indicate perceiving the stimulus as sounding more like normal speech.

Methods

Participants:

In this experiment, we established a "stopping rule," whereby data collection ceased when we had collected data from 60 participants or the semester ended. By the end of the semester we collected data from 58 native English speakers who were recruited from a pool of students enrolled in Introductory Psychology at the University of Kansas. Participants received partial credit towards the completion of the course for their participation. All were native English speakers, and none reported a hearing or speech disorder. Written informed consent was obtained before participating in the experiment, and this experiment was approved by the institutional review board at the University of Kansas.

Stimuli:

Thirty-six high clustering coefficient words and 36 low clustering coefficient words originally used in Chan and Vitevitch (2009) were used as stimuli in this experiment. All words were produced by a native speaker of American English (M.S.V.) speaking at a normal rate and loudness and recorded on high-quality recording equipment as described in Chan and Vitevitch (2009). The pronunciation of each word was verified for correctness and minimal intonation changes.

The high clustering coefficient words were randomly assigned to 9 groups of four words such that each word was only used once (e.g., full-bug-leap-mouse). The low clustering coefficient words were then assigned to nine groups of four words such that the onsets of the words in each list matched the onsets of the words in the lists containing high clustering coefficient words (e.g., fell-beat-ledge-mile). See the Supplementary Material for the words used in this experiment.

Sound files were concatenated together to form lists using the open-source Audacity 2.2.2 software. In all lists, words were separated by approximately 25 ms of silence, and no additional silence was included at the beginning nor the end of each sound file. Further information on the lexical characteristics of these words was originally reported in Chan and Vitevitch (2009), but information on the lexical characteristics most relevant to the present experiment is repeated below.

Praat (Boersman & Weenink, 2020; version 6.1.15) was used to confirm that the minimum pitch, maximum pitch, and mean pitch of the word lists did not differ between the two

conditions. For the minimum pitch, the lists with high clustering coefficient words (M = 85.37 Hz; SD = 4.2) did not differ from the lists with low clustering coefficient words, M = 88.36 Hz; SD = 3.69; $t(16) = 1.61$, $p = .13$. For the maximum pitch, the lists with high clustering coefficient words (M = 243.33 Hz; SD = 158.92) did not differ from the lists with low clustering coefficient words, M = 208.41 Hz; SD = 146.73; $t(16) = 0.48$, $p = .63$. For the mean pitch, the lists with high clustering coefficient words (M = 111.22 Hz; SD = 5.57) did not differ from the lists with low clustering coefficient words, M = 110.02 Hz; SD = 2.86; $t(16) = 0.58$, $p = .57$.

Clustering coefficient:

Clustering coefficient measures the probability that the neighbours of a given node are also neighbours of each other and has a range from 0 to 1. A clustering coefficient of 1 means every neighbour is interconnected, whereas a clustering coefficient of 0 means no neighbours are connected with each other. Clustering coefficient for a node (which represent words in a lexical network) is calculated by dividing the number of connections among neighbours by the total possible number of connections if all neighbours were interconnected. For additional analyses of the clustering coefficient and how it is related to other lexical measures, such as neighbourhood density, see Vitevitch et al. (2012).

The high clustering coefficient words had a mean clustering coefficient value of 0.171 (SD = 0.02; $n = 36$) and the low clustering coefficient words had a mean clustering coefficient value of 0.122 (SD = 0.01; $n = 36$). Using a two-tailed, independent samples t-test, we found the difference in clustering coefficient to be statistically significant, $t(70) = 13.11$, $p < .0001$. Despite differing in clustering coefficient, the two groups of words were equivalent on other measures including subjective familiarity, word frequency, neighbourhood density, neighbourhood frequency, and phonotactic probability.

Subjective Familiarity:

Subjective familiarity was measured on a 7-point scale (Nusbaum et al., 1984). Higher clustering coefficient words had a mean familiarity value of 6.91 (SD = 0.18), and lower clustering coefficient words had a mean familiarity value of 6.95 (SD = 0.09), $t(70) = 1.13$, $p = .19$. All words, regardless of clustering coefficient, were therefore highly familiar words.

Word Frequency:

Word frequency refers to how often a word occurs in the language. Mean log word frequency ($\log_{10}$ of the raw values from Kucera & Francis, 1967) was 2.38 (SD = 0.72) for the high clustering coefficient words and 2.41 (SD = 0.61) for the low clustering coefficient words, $t(70) = 0.23$, $p = .82$. This indicates that the two conditions contained words that occurred equally often in the language.

Neighbourhood Density:

Neighbourhood density is the number of words that are neighbours to the target word. A word was considered a neighbour of a target word if the substitution, deletion, or addition of a single phoneme transformed a word into the target word. For example, the word *pin* has phonological neighbour words such as _in, *s*pin, *f*in, p*u*n, and pi*t*. This is a commonly used

metric to assess phonological similarity (Greenberg & Jenkins, 1967; Landauer & Streeter, 1973; Luce & Pisoni, 1998). The neighbourhood density value for high clustering coefficient words was 21.03 neighbours (SD = 5.68); low clustering coefficient words had a neighbourhood density value of 21.94 neighbours (SD = 7.04), $t(70) = 0.61$, $p = .55$. Thus, the two conditions contained words that had approximately the same number of phonological neighbours.

Neighbourhood Frequency:

Neighbourhood frequency refers to how often the neighbours of a given word occur in the language. Mean neighbourhood frequency ($\log_{10}$ of the raw values from Kucera & Francis, 1967) was 2.01 (SD = 0.21) for the high clustering coefficient words and 2.03 (SD = .20) for the low clustering coefficient words, $t(70) = 0.44$, $p = .66$. This suggests that the two lists contained words with comparable neighbourhood frequency.

Phonotactic Probability:

Phonotactic probability measures how often a certain segment occurs in a certain position in a word (positional segment frequency) and the segment-to-segment co-occurrence probability (biphone frequency; Vitevitch & Luce, 1998, 2005) We obtained these values from the Web-based calculator described in Vitevitch and Luce (2004). The mean positional segment frequency for high clustering coefficient words was .1379 (SD = 0.03), and 0.143 (SD = 0.04) for low clustering coefficient words, $t(70) = 0.56$, $p = .58$. The mean biphone frequency for both high and clustering coefficient words was 0.006 (SD = 0.004), $t(70) = 0.46$, $p = .65$. This suggests that the two lists contained words with comparable phonotactic probability.

Procedure:

Participants were tested individually. Each participant was seated in front of an iMac computer running PsyScope 1.2.2 (Cohen et al., 1993). This programme controlled stimulus presentation and collected responses.

The word “READY” appeared on the computer screen for 500 ms at the start of each trial. Participants then heard one of the randomly selected word lists repeated 10 times through a set of Beyerdynamic DT 100 headphones at a comfortable listening level. After the repetitions, participants were instructed to use the number pad on the keyboard to rate the list on a scale of 1 (sounded more like speech) to 5 (sounded more like song). Each trial was only presented once, but participants were allowed as much time as they needed to respond. In total, the experiment lasted approximately 10 to 15 min.

Note that in some studies of the speech-to-song illusion, participants rate the stimulus after it has been presented only once, and those ratings are compared to the ratings after the stimulus is played several times to demonstrate that repetition of the phrase leads to the speech-to-song illusion (as indicated by increases in song-like ratings to the final repetition compared to the initial repetition). Although this method has sometimes been employed—as in Experiment 1 (but not Experiment 2) of Deutsch et al. (2011), and in Experiments 1 and 3 (but not Experiments 2, 4, 5, or 6) of Castro et al. (2018)—we did not employ it in the

present study because as the numerous studies described in the introduction indicate, the speech-to-song phenomenon is very well established, having been replicated in a number of laboratories around the world. The preponderance of evidence clearly indicates that the transformation of a repeated speech stimulus to a song-like percept is a genuine phenomenon and is not in question.

Furthermore, both theories being tested in the present study—NST and the repetition account—predict that the speech-to-song illusion will be elicited. In the present study, the important difference in ratings is the potential difference that might be observed between the lists containing words with low clustering coefficient and the lists containing words with high clustering coefficient. The repetition account does not predict that there should be a difference between the two conditions, whereas NST does predict a difference between the two conditions. We, therefore, decided to keep the procedure focused on that crucial difference.

Results

All ratings from the 58 participants were used in the analysis reported in the following. High clustering coefficient words received a mean rating of 2.24 ($SD = 0.64$), whereas low clustering coefficient words received a mean rating of 2.78 ($SD = 0.58$). The difference between ratings was statistically significant, $t(57) = 6.60$, $p < .0001$, and the size of the effect was considered large (Cohen's $d = .89$ as computed by https://webpower.psychstat.org/models/means01/effectsize.php). As predicted, low clustering coefficient words were perceived as more song-like than high clustering coefficient words.

Discussion

In the present experiment, we used a measure of phonological similarity among words derived from network science—namely, the clustering coefficient (Chan & Vitevitch, 2009)—to further test the mechanisms of priming, activation, and satiation found in NST as a way to account for the S2S illusion. We predicted that high and low C words would be primed, activated, and satiated, thereby eliciting a song-like percept. We further predicted that there would be a difference in how quickly words with high versus low C recovered from satiation, which would affect the extent to which high versus low C words would be perceived as song-like. Indeed, the results of the present experiment showed that low clustering coefficient words were perceived as more song-like than high clustering coefficient words, as predicted by NST.

Reasoning from the previous psycholinguistic studies that examined the influence of C on various language and memory processes (Chan & Vitevitch, 2009, 2010; Goldstein & Vitevitch, 2014; Vitevitch et al., 2012), as well as from the computer simulations by Vitevitch et al. (2011) and Siew (2019), we predicted that the small number of interconnections among the neighbours for words with low C would result in some priming from the neighbours being transmitted back to the target word, but most of the priming being transmitted to the rest of the network. With less priming accumulating over time and across connections words with low C would stay satiated longer than high C words, allowing

priming to continue to affect the syllable nodes (which carry the rhythmic information of language). With sustained priming of the syllable nodes, the song-like percept could be maintained, resulting in the higher song-like ratings that were observed.

In contrast, for high *C* words, priming is "trapped" among the highly interconnected neighbours and transmitted back to the target word rather than being dispersed to the rest of the network as happens for low *C* words. The trapped priming would accumulate more quickly in the lexical node enabling words with high *C* to recover from satiation and be activated again more quickly than a word with low *C*. The activation of lexical nodes for high *C* words would again bring the speech percept to awareness, thereby weakening the song-like percept. This would result in lower song-like ratings for word lists containing words with high *C* compared to word lists containing words with low *C*, as was observed.

The result of the present experiment provides evidence to further support the mechanisms found in NST—priming, activation, and satiation—as an explanation for the S2S illusion as proposed by Castro et al. (2018). It is unclear how accounts of the S2S illusion that appeal only to repetition (Margulis, 2013; Margulis & Simchy-Gross, 2016; Rowland et al., 2019) would explain the results of the present experiment. Repetition theories of the S2S illusion would predict that the repeated word lists would elicit the S2S illusion, as was observed in the present study. However, such approaches would not predict (nor can they account for) the observed difference between the word lists varying in clustering coefficient.

Experiment 2

In Experiment 1, we examined how the mechanisms found in NST—priming, activation, and satiation—were differentially influenced by another measure of phonological similarity known as the clustering coefficient. Because repetition theories of the S2S illusion cannot account for the differential influence of that phonological variable on the illusion, the result of Experiment 1 provided additional support for the mechanisms found in NST as an explanation for the S2S illusion (Castro et al., 2018).

In the present experiment, we wished to further test how the mechanisms found in NST—priming, activation, and satiation—might influence the S2S illusion. Rather than focus on the mechanisms of priming and activation as in Experiment 1, we wished in the present experiment to focus on the mechanism of satiation, which was not tested as directly as priming and activation in Castro et al. (2018). Recall that when lexical nodes are satiated, the higher level information associated with that node (such as semantic information) is no longer available to conscious awareness (see Figure 1). Rather than manipulate another phonological characteristic of the repeated words to produce a differential influence as in Experiment 1, we instead in the present experiment manipulated a semantic characteristic of the repeated words. In further contrast to Experiment 1, in the present experiment, we predicted that because the lexical nodes are satiated and the higher level information associated with that node is no longer available to conscious awareness, that there should be *no difference* in song-like ratings for lists of words that differ in a semantic variable.

To test this unique prediction, we manipulated the semantic characteristic known as *emotional arousal* (as measured by the Affective Norms for English Words [*ANEW*] database; Bradley & Lang, 1999). Of the various semantic features that could be manipulated, we selected emotional arousal because relationships among emotion, music, and language have been widely studied (e.g., Asaridou & McQueen, 2013; Bigand et al., 2005; Hernández et al., 2019; Koelsch et al., 2006 Margulis, 2013; Martin-Loeches et al., 2012; Patel, 2008; Tay & Ng, 2019). Furthermore, the ANEW database (Bradley & Lang, 1999) is widely used and has been adapted to a number of other languages (e.g., *German*: Schmidtke et al., 2014; *Italian*: Montefinese et al., 2014; *European Portuguese*: Soares et al., 2012; *Spanish*: Redondo et al., 2007). Furthermore, the emotional information of words has been shown to influence reaction times in various language processes, including semantic categorization tasks (Newcombe et al., 2012), word naming (Moffat et al., 2015), colour naming performance in the Stroop task (Siakaluk et al., 2014), and the lexical decision task (Siakaluk et al., 2016). It, therefore, seemed reasonable to consider the specific semantic variable of emotional arousal (see also Kuperman et al., 2014) in the context of the S2S illusion.

Given that words with higher emotional experience ratings tend to be responded to more quickly than words with lower emotional experience ratings (e.g., Moffat et al., 2015; Siakaluk et al., 2016), it further seemed logical for repetition theories of the S2S illusion to predict that in a S2S illusion task lists containing words with meanings that are emotionally arousing (e.g., *passion, killer, rage, startled*) would evoke higher song-like ratings compared to words with meanings that are less emotionally arousing (as measured by the Affective Norms for English Words [ANEW] database; Bradley & Lang, 1999). Manipulating emotional arousal gave us the opportunity to formulate a prediction that contrasted with the very different prediction made by NST, namely no difference in song-like ratings because semantic information is not accessible when lexical nodes are satiated.

Alternatively, an anonymous reviewer suggested the possibility that arousal-based competition theory (Mather & Sutherland, 2011) would predict that to avoid deception attention would be allocated to the highly arousing words used in the S2S illusion task. This would instead result in lists containing words with meanings that are more emotionally arousing to evoke lower song-like ratings compared to words with meanings that are less emotionally arousing. Having two theoretically grounded predictions for differences in the song-like ratings (although in opposite directions) make for an even stronger test of the prediction derived from NST, where satiation of the lexical nodes would result in the inability to access semantic/emotional information, and therefore no difference in the song-like ratings.

Predicting equivalence between two conditions, as we have done in this case, is problematic for traditional statistical tests, such as the *t*-test used in Experiment 1, that are designed to reject the null hypothesis. In a traditional *t*-test, non-significant values indicate that the evidence is not strong enough to suggest a difference between the two distributions. Such results mean that equality of the two distributions cannot be ruled out (i.e., there is no effect), but such results could also be due to a lack of power. Unfortunately, there is no way

to distinguish between those two possibilities, making a non-significant null-hypothesis test problematic to interpret.

To provide evidence that two groups are indeed equivalent or that the difference between them is so small as to be inconsequential, a test of equivalence such as the *two one-sided tests* (*TOST*) can be used (García-Pérez & Alcalá-Quintana, 2011; Lakens et al., 2018; Rogers et al., 1993; Rose et al., 2018; Walker & Nowacki, 2010). Conceptually, the TOST reverses the null and alternative hypotheses found in traditional null-hypothesis testing. In TOST, the null hypothesis states that the two distributions are not equivalent, and the alternative hypothesis states that the two distributions are equivalent if the difference between the two distributions falls within a preset range.

Equivalence testing using methods such as TOST are common in pharmaceutical research, where a researcher may wish to demonstrate that the generic version of a drug produces similar effects as the brand-name version of the drug (Food and Drug Administration, 2001). In other areas, such as Psychology (Lakens et al., 2018; Rogers et al., 1993) and Ecology (Rose et al., 2018), equivalence testing has taken a bit longer to find widespread use. Perhaps using this method to test the prediction derived from NST of equivalence in the S2S illusion task for word lists varying in a semantic characteristic (i.e., emotional arousal) will demonstrate the value of equivalence testing to a wider range of Psychologists.

Methods

Participants:

As in Experiment 1, we established a “stopping rule,” whereby data collection ceased when we had collected data from 60 participants or the semester ended. Due to a number of factors (e.g., when data collection started in the semester, the number of competing experiments available for participants to engage in, scheduling-related issues, etc.), data from only 40 participants were collected when the end of the semester was reached. All of these participants were English speakers (none reported a hearing or speech disorder) that were recruited from a pool of students enrolled in Introductory Psychology at the University of Kansas. Participants received partial credit towards the completion of the course for their participation. Written informed consent was obtained before participating in the experiment, and this experiment was approved by the institutional review board at the University of Kansas. Of the 40 participants, one did not follow instructions, and two did not have their data recorded due to a technical error, leaving data from 37 participants to be used in our analyses.

Stimuli:

Fifty-six English words were selected from the Affective Norms for English Words (ANEW) database (Bradley & Lang, 1999). All words were produced by a native speaker of American English (M.S.V.) speaking at a normal rate and loudness and recorded on the same equipment as used in Experiment 1. The pronunciation of each word was verified for correctness and minimal intonation changes.

The sound files containing individual words were then concatenated into 7 lists containing 4 words with high arousal and 7 lists containing 4 words with low arousal. The same procedure used in Experiment 1 to match the onsets of the words across lists was used in the present experiment. In addition, each list contained words with the same number of syllables. See the Supplementary Material for the words used in this experiment.

Praat (Boersman & Weenink, 2020; version 6.1.15) was used to confirm that the minimum pitch, maximum pitch, and mean pitch of the word lists containing the sound files did not differ between the two conditions. For the minimum pitch, the lists with high arousal words (M = 85.04 Hz; SD = 2.16) did not differ from the lists with low arousal words, M = 84.88 Hz; SD = 3.25; $t(12) = 0.11$, $p = .92$. For the maximum pitch, the lists with high arousal words (M = 318.29 Hz; SD = 174.27) did not differ from the lists with low arousal words, M = 269.54 Hz; SD = 167.28; $t(12) = 0.53$, $p = .60$. For the mean pitch, the lists with high arousal words (M = 115.77 Hz; SD = 9.13) did not differ from the lists with low arousal words, M = 113.61 Hz; SD = 5.46; $t(12) = 0.53$, $p = .6$.

As described in Bradley and Lang (1999), *arousal* was measured on a 9-point rating scale. Words in the high arousal condition had a mean rating = 6.61 (SD = .66), and words in the low arousal condition had a mean rating = 3.92 (SD = .53). An unpaired two-tailed t-test showed a statistically significant difference in the ratings for the two arousal conditions, $t(54) = 16.79$, $p < .0001$.

Importantly, word frequency (as measured by Kucera & Francis, 1967 as used in Bradley & Lang, 1999) did not differ between the two conditions, as shown in an unpaired two-tailed t-test, $t(54) = 0.09$, $p = .925$. Words in the high arousal condition had a mean frequency of occurrence = 32.96 (SD = 33.37), and words in the low arousal condition had a mean frequency of occurrence = 31.86 (SD = 51.80).

Procedure:

The same equipment and procedure used in Experiment 1 were used in the present experiment. The only exception is the number of word lists used in each experiment. As in Experiment 1, we elected not to have participants rate the stimuli after one presentation and again after several presentations because both theories being tested in the present study—NST and the repetition account—predict that the speech-to-song illusion will be elicited. In the present study, the important difference in ratings is the potential difference that might be observed between the lists containing words with low arousal and the lists containing words with high arousal. The repetition account predicts that there should be a difference between the two conditions, whereas NST predicts that there will not be a difference between the two conditions. We, therefore, decided to keep the procedure focused on that crucial difference.

Results

As recommended by Lakens et al. (2018), we report both a traditional null-hypothesis significance test and an equivalence test using the two one-sided tests procedure as implemented on the Excel spreadsheet described in Lakens (2017). A traditional null-hypothesis significance test using a paired t-test was used to compare song-like ratings for

the High and Low-Arousal lists. For the High-Arousal condition the *mean* rating = 2.46 (*SD* = 0.70), and for the Low-Arousal condition the *mean* rating = 2.54 (*SD* = 0.62). This difference was not statistically significant, $t(36) = 1.05$, $p = .2988$.

To determine whether the mean ratings for the High and Low Arousal conditions in the S2S illusion task were statistically equivalent (as predicted by NST), we performed the TOST procedure for dependent samples with equivalence bounds based on raw scores. We provide information on this procedure directly from Lakens et al. (2018, p. 260):

> In the TOST procedure, the first one-sided test is used to test the estimate against values at least as extreme as the lower equivalence bound (Δ_L) and the second one-sided test is used to test the estimate against values at least as extreme as the upper equivalence bound (Δ_U). Even though the TOST procedure consists of two one-sided tests, it is not necessary to control for multiple comparisons because both tests need to be statistically significant for the researcher to draw a conclusion of statistical equivalence. Consequently, when reporting an equivalence test, it suffices to report the one-sided test with the smaller test statistic (e.g., *t*) and thus the larger *p* value. A conclusion of statistical equivalence is warranted when the larger of the two *p* values is smaller than alpha. If the observed effect is neither statistically different from zero nor statistically equivalent, there is insufficient data to draw conclusions.

The smallest effect size of interest (*SESOI*) used to construct the upper and lower equivalence bounds was set in terms of a raw mean difference; specifically, 0.5 points on the 5-point rating scale used in many studies of the S2S illusion. Using raw scores to estimate effect size rather than standardised effect measures such as Cohen's *d* allows us to more easily interpret the effect size because we maintain the original units of measure (Baguley, 2009). Furthermore, the five-point scale used in many studies of the S2S illusion was used in Experiment 1 where a difference of approximately 0.5 rating units was observed in the statistically significant effect that was reported. Furthermore, the size of the effect observed in Experiment 1 was comparable to the size of the effects observed in the six experiments reported in Castro et al. (2018).

The TOST procedure indicated that the observed effect size ($d_z = -0.17$) was significantly within the equivalent bounds of −0.5 and 0.5 scale points (or in Cohen's d_z: −1.04 and 1.04), $t(36) = 5.33$, $p < .0001$. The results of the TOST procedure indicate that the difference between the High and Low Arousal conditions in the S2S illusion task is smaller than what is considered meaningful and statistically falls within the interval indicated by the equivalence bounds. That is, the ratings to High Arousal words were statistically equivalent to the ratings to the Low Arousal words, as predicted by NST.

We wished to bring attention in Psychology to equivalence testing by using the null-hypothesis version of equivalence testing (i.e., the TOST method). However, Bayesian versions of equivalence testing also exist. Given that Bayesian statistical analysis is increasingly being used in Psychology (e.g., Etz & Vandekerckhove, 2018), we include the results of a Bayesian equivalence test for comparison.

JASP (JASP Team, 2020) was used to compute a Bayesian Equivalence *t*-test for paired samples. The overlapping-hypothesis Bayes factor ($BF_{\notin\in}$) was 0.002 and the nonoverlapping-hypothesis Bayes factor ($BF_{\in\notin}$) was 467, which provide strong (Raftery, 1995) to decisive (Jeffreys, 1961) evidence that the ratings to High Arousal words are equivalent to the ratings to the Low Arousal words, as predicted by NST.

Discussion

Given that previous work in psycholinguistics found that words with higher emotional experience ratings were responded to more quickly than words with lower emotional experience ratings (e.g., Moffat et al., 2015; Siakaluk et al., 2016), it was logical to infer that repetition theories of the S2S illusion would predict that lists containing words with meanings that are emotionally arousing (Bradley & Lang, 1999) would evoke higher song-like ratings compared to words with meanings that are less emotionally arousing. Alternatively, arousal-based competition theory (Mather & Sutherland, 2011) would predict that to avoid deception attention would be allocated to the highly arousing words, resulting in lists containing words with meanings that are more emotionally arousing evoking lower song-like ratings compared to words with meanings that are less emotionally arousing.

In contrast, the mechanisms in NST that underlie the S2S illusion—priming, activation, and satiation—made a different prediction regarding the influence of semantic information on the S2S illusion. Specifically, lexical nodes become satiated as the stimulus repeats, leading to the initial speech percept being lost, and to the emergence of the song percept because the repeating stimulus continues to prime the syllable nodes (where the rhythmic information of language is represented). Importantly, satiation of the lexical nodes in NST not only leads to loss of the speech percept, but also to the loss of the higher-level information connected to the lexical node (refer back to Figure 1), such as the semantic information and emotional arousal associated with the meaning of the word. NST therefore predicted no difference in song-like ratings for lists of words with high or low arousal in an S2S illusion task.

The song-like ratings observed in the present experiment were comparable in magnitude to the song-like ratings observed in Experiment 1, suggesting that the speech to song illusion was elicited for these lists of words. Crucially, however, the difference in the ratings for the high and low arousal conditions were found to be statistically equivalent using the TOST procedure. The equivalent rating for high and low arousal word lists is consistent with the prediction derived from NST and contrasts with the prediction derived from the repetition theory of the S2S illusion.

General discussion

Although repetition of the stimulus is clearly necessary to elicit the S2S illusion, repetition by itself is not sufficient to explain various aspects of the S2S illusion, nor the findings from the two experiments reported. The results obtained in the present studies provide additional evidence for the account of the speech to song illusion offered by NST. In NST, the lexical nodes are initially activated giving rise to the speech percept. The continued presentation of the stimulus results in the lexical nodes satiating, which means the speech percept and all

higher level information connected to the lexical node is temporarily unavailable. The repeated presentation of the stimulus continues to prime the syllable nodes, which contain the rhythmic information of language, and gives rise to the song-like percept.

In Experiment 1, we used words varying in the clustering coefficient, a measure derived from network science, to manipulate the amount of priming that would be transmitted to a lexical node after it was satiated. Words with high clustering coefficient, or phonological neighbours that were also neighbours of each other, “trapped” priming near the lexical node, enabling it to more quickly accumulate sufficient amounts of priming to again activate the lexical node compared to words with low clustering coefficient, or phonological neighbours that tended not to be neighbours of each other. When words with high clustering coefficient are again activated the speech percept returns, thereby leading to lower song-like ratings for words with high clustering coefficient.

For words with low clustering coefficient, once the lexical node has satiated priming will be dispersed to the rest of the network (as demonstrated in computer simulations by Vitevitch et al., 2011 and Siew, 2019), which keeps the lexical node satiated for a longer time. With the lexical node satiated, the repeating stimulus will continue to prime the syllable nodes and maintain the song-like percept. Although this effect was predicted by NST, it is unclear how an account of the speech to song illusion that argues only for the role of repetition would account for the differential ratings to word lists varying in phonological clustering coefficient.

We sought in Experiment 2 to focus our test of the NST account of the S2S illusion on the mechanism of satiation and the role it plays in the S2S illusion. Although priming and activation had been examined extensively in Castro et al. (2018), the mechanism of satiation received less direct examination in that previous study. In NST, when a lexical node is satiated after being repeatedly activated, higher level information connected to that lexical node can no longer be accessed. Therefore, words varying in emotional arousal should not affect song-like ratings in a speech to song illusion task, in contrast to reasonable predictions derived from the well-studied influences of emotion on music and language and previous psycholinguistic studies, and reasonable predictions derived from arousal-based competition theory (Mather & Sutherland, 2011). Using equivalence testing (i.e., the TOST procedure and a Bayesian version)—a statistical analysis that is used widely in other fields, but less so in Psychology—we demonstrated that the song-like ratings to word lists that were high or low in emotional arousal were indeed statistically equivalent, as predicted by NST.

By examining the speech to song illusion in the context of NST, we place the illusion into a rich theoretical context that allows us to connect this illusion to a wide range of perceptual and cognitive phenomena related to music and language processing more broadly. For example, music is used in many therapeutic interventions for speech and language disorders (e.g., Cohen, 1994; Kasdan & Kiran, 2018). Thus, the insight gained from continued study of the speech to song illusion could not only have important implications for increasing our understanding of the perceptual and cognitive systems that underlie the illusion, but may also lead to the future development of novel interventions for certain speech- and language-related disorders.

Continued research on the speech-to-song illusion may also help determine whether the mechanisms of priming, activation, and satiation found in NST can also account for another auditory illusion that—on the surface—resembles the speech-to-song illusion, namely the sound-to-music illusion (Margulis & Simchy-Gross, 2016; Rowland et al., 2019; Simchy-Gross & Margulis, 2018; Tierney et al., 2018). In the sound-to-music illusion, complex tones or environmental sounds (*e.g.*, ice cracking) are repeated, and begin to take on a musical quality. As we have demonstrated in the speech-to-song illusion, repetition of the stimulus is probably necessary to elicit the sound-to-music illusion, but repetition alone is unlikely to be sufficient for explaining how or why the sound-to-music illusion occurs.

Although we have not closely or directly examined the sound-to-music illusion, we believe it is unlikely that the mechanisms of priming, activation, and satiation as found in NST would be able to satisfactorily account for the sound-to-music illusion simply because it is unlikely that any nodes in NST, a model of language perception and production, would be primed, activated, and satiated by the incoming acoustic, but non-speech, stimulus typically used to elicit the sound-to-music illusion (*cf.*, Bartolotti et al., 2020; Koranda et al., 2020). It is possible, however, that mechanisms like priming, activation, and satiation at another level of processing (i.e., not cortical) might account for the sound-to-music illusion. We raise this possibility of redundant processes by analogy to the various mechanisms involved in colour perception, including different types of cones and the colour-opponent mechanism found in the ganglia in the retina, in addition to cortical areas being involved in the perception of colour (Solomon & Lennie, 2007). We further observe that the ability to synchronise body movement to an external beat (i.e., rhythmic entrainment) has evolved in a wide range of species, including fireflies, birds, dolphins, and humans (Wilson & Cook, 2016), further suggesting that multiple mechanisms at different neurological levels could be involved in the transformation of auditory stimuli of various types into a music-like percept. Admittedly, these concluding statements are speculative. Therefore, we eagerly await the findings of future tests of the NST and other accounts of the speech-to-song and sound-to-music illusions.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We wish to thank the Initiative for Maximising Student Development (IMSD) at the University of Kansas for their financial support of J.W.N. while working on Experiment 1. We wish to thank the Centre for Undergraduate Research at the University of Kansas for their financial support of E.H. while working on Experiment 2, which also served to partially fulfil the requirements for Departmental Honours in Psychology at the University of Kansas.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

Asaridou SS, & McQueen JM (2013). Speech and music shape the listening brain: Evidence for shared domain-general mechanisms. Frontiers in Psychology, 4, 321. [PubMed: 23761776]

Baguley T. (2009). Standardized or simple effect size: What should be reported? British Journal of Psychology, 100, 603–617. [PubMed: 19017432]

Bartolotti J, Schroeder SR, Hayakawa S, Rochanavibhata S, Chen P, & Marian V. (2020). Listening to speech and non-speech sounds activates phonological and semantic knowledge differently. Quarterly Journal of Experimental Psychology, 73, 1135–1149.

Bashford JA Jr., Warren RM, & Lenz PW (2006). Polling the effective neighborhoods of spoken words with the verbal transformation effect. Journal of the Acoustical Society of America Express Letters, 119, EL55–EL59.

Bigand E, Filipic S, & Lalitte P. (2005). The time course of emotional responses to music. Annals of the New York Academy of Science, 1060, 429–437.

Boersman P, & Weenink D. (2020). Praat: Doing phonetics by computer [Computer program, Version 6.1.15]. http://www.praat.org

Boyette L-L, Isvoranu AM, Schirmbeck F, Velthorst E, Simons CJP, Barrantes-Vidal N, Bressan R, Kempton MJ, Krebs M-O, McGuire P, Nelson B, Nordentoft M, Riecher-Rössler A, Ruhrmann S, Rutten BP, Sachs G, Valmaggia LR, van der Gaag M, BorsboomD de Haan L, & van Os J. (2020). From speech illusions to onset of psychotic disorder: Applying network analysis to an experimental measure of aberrant experiences. Schizophrenia Bulletin Open, 1, Sgaa025. 10.1093/schizbullopen/sgaa025

Bradley MM, & Lang PJ (1999). Affective norms for English words (ANEW): Instruction manual and affective ratings [Technical Report C-1]. The Center for Research in Psychophysiology, University of Florida.

Burke DM, MacKay DG, Worthley JS, & Wade E. (1991). On the tip-of-the-tongue: What causes word finding failures in young and older adults? Journal of Memory and Language, 30, 542–579.

Castro N, Mendoza JM, Tampke EC, & Vitevitch MS (2018). An account of the speech-to-song illusion using node structure theory. PLOS ONE, 13(6), Article e0198656. 10.1371/journal.pone.0198656

Chan KY, & Vitevitch MS (2009). The influence of the phonological neighborhood clustering coefficient on spoken word recognition. Journal of Experimental Psychology: Human Perception and Performance, 35, 1934–1949. [PubMed: 19968444]

Chan KY, & Vitevitch MS (2010). Network structure influences speech production. Cognitive Science, 34, 685–697. [PubMed: 21564230]

Cohen J, MacWhinney B, Flatt M, & Provost J. (1993). PsyScope: An interactive graphic system for defining and controlling experiments in the psychology laboratory using Macintosh computers. Behavior Research, Methods, Instruments, and Computers, 25, 257–271.

Cohen NS (1994). Speech and song: Implications for therapy. Music Therapy Perspectives, 12(1), 8–14.

Cutler A. (1991). Linguistic rhythm and speech segmentation. In Sundberg J, Nord L, Carlson R (eds.), Wenner-Gren center international symposium series: Music, language, speech and brain (pp. 157–166). Palgrave.

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

Etz A, & Vandekerckhove J. (2018). Introduction to Bayesian inference for psychology. Psychonomic Bulletin & Review, 25, 5–34. [PubMed: 28378250]

Falk S, & Rathcke T. (2010). On the speech-to-song illusion: Evidence from German. Speech Prosody, 100169, 1–4.

Food and Drug Administration. (2001). Guidance for industry: Statistical approaches to establishing bioequivalence. Center for Drug Evaluation and Research, U.S. Food and Drug Administration. https://www.fda.gov/downloads/drugs/guidances/ucm070244.pdf

Fujii S, & Wan CY (2014). The role of rhythm in speech and language rehabilitation: The SEP hypothesis. Frontiers in Human Neuroscience, 8, 777. 10.3389/fnhum.2014.00777 [PubMed: 25352796]

García-Pérez MA, & Alcalá-Quintana R. (2011). Testing equivalence with repeated measures: Tests of the difference model of two-alternative forced-choice performance. The Spanish Journal of Psychology, 14, 1023–1049. [PubMed: 22059346]

Goldstein R, & Vitevitch MS (2014). The influence of clustering coefficient on word-learning: How groups of similar sounding words facilitate acquisition. Frontiers in Language Sciences, 5, 01307.

Greenberg JH, & Jenkins JJ (1967). Studies in the psychological correlates of the sound system of American English. In Jakobovits LA & Miron MS (Eds.), Readings in the psychology of language (pp. xi, 636, 186–200). Prentice Hall.

Gregory RL (1968). Perceptual illusions and brain models. Proceedings of the Royal Society of London. Series B, Biological Sciences, 171, 279–296.

Hernández M, Palomar-García MÁ, Nohales-Nieto B, Olcina-Sempere G, Villar-Rodríguez E, Pastor R, Ávila C, & Parcet MA (2019). Separate contribution of striatum volume and pitch discrimination to individual differences in music reward. Psychological Science, 30, 1352–1361. [PubMed: 31340130]

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

Team JASP. (2020). JASP (Version 0.13.1) [Computer software].

Jeffreys H. (1961). Theory of probability (3rd ed.). Oxford University Press.

Kasdan A, & Kiran S. (2018). Please don't stop the music: Song completion in patients with aphasia. Journal of Communication Disorders, 75, 72–86. 10.1016/j.jcomdis.2018.06.005 [PubMed: 30031236]

Koelsch S, Fritz T, Cramon DY, Muller K, & Friederici AD (2006). Investigating emotion with music: An fMRI study. Human Brain Mapping, 27, 239–250. [PubMed: 16078183]

Koranda MJ, Bulgarelli F, Weiss DJ, & MacDonald MC (2020). Is language production planning emergent from action planning? A preliminary investigation. Frontiers in Psychology, 11, 1193. 10.3389/fpsyg.2020.01193 [PubMed: 32581969]

Kucera H, & Francis WN (1967). Computational analysis of present day American English. Brown University Press.

Kuperman V, Estes Z, Brysbaert M, & Warriner AB (2014). Emotion and language: Valence and arousal affect word recognition. Journal of Experimental Psychology: General, 143, 1065–1081. [PubMed: 24490848]

Lakens D. (2017). Equivalence tests: A practical primer for t tests, correlations, and meta-analyses. Social Psychological and Personality Science, 8, 355–362. [PubMed: 28736600]

Lakens D, Scheel AM, & Isager PM (2018). Equivalence testing for psychological research: A tutorial. Advances in Methods and Practices in Psychological Science, 1, 259–269.

Lambert WE, & Jakobovits LA (1960). Verbal satiation and changes in the intensity of meaning. Journal of Experimental Psychology, 60, 376–383. [PubMed: 13758466]

Landauer TK, & Streeter LA (1973). Structural differences between common and rare words: Failure of equivalence assumptions for theories of word recognition. Journal of Verbal Learning and Verbal Behavior, 12, 119–131.

Luce PA, & Pisoni DB (1998). Recognizing spoken words: The neighborhood activation model. Ear and Hearing, 19, 1–36. [PubMed: 9504270]

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

MacKay DG, & Burke DM (1990). Cognition and aging: A theory of new learning and the use of old connections. In Hess T (Ed.), Aging and cognition: Knowledge organization and utilization (pp. 213–263). Elsevier.

MacKay DG, Stewart R, & Burke DM (1998). H. M.'s language production deficits: Implications for relations between memory, semantic binding, and the hippocampal system. Journal of Memory and Language, 38, 28–69.

MacKay DG, 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.

Margulis EH (2013). Repetition and emotive communication in music versus speech. Frontiers in Psychology, 4, 167. [PubMed: 23576998]

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

Martin-Loeches M, Fernandez A, Schacht A, Sommer W, Casado P, Jimenez-Ortega L, & Fondevila S. (2012). The influence of emotional words on sentence processing: Electrophysiological and behavioral evidence. Neuropsychologia, 50, 3262–3272. [PubMed: 22982604]

Mather M, & Sutherland MR (2011). Arousal-based competition in perception and memory. Perspectives on Psychological Science, 6, 114–133. [PubMed: 21660127]

Moffat M, Siakaluk PD, Sidhu DM, & Pexman PM (2015). Situated conceptualization and semantic processing: Effects of emotional experience and context availability in semantic categorization and naming tasks. Psychonomic Bulletin & Review, 22, 408–419. 10.3758/s13423-014-0696-0 [PubMed: 25092388]

Montefinese M, Ambrosini E, Fairfield B, & Mammarella N. (2014). The adaptation of the Affective Norms for English Words (ANEW) for Italian. Behavior Research Methods, 46, 887–903. [PubMed: 24150921]

Newcombe PI, Campbell C, Siakaluk PD, & Pexman PM (2012). Effects of emotional and sensorimotor knowledge in semantic processing of concrete and abstract nouns. Frontiers in Human Neuroscience, 6, 275. 10.3389/fnhum.2012.00275 [PubMed: 23060778]

Nusbaum HC, Pisoni DB, & Davis CK (1984). Sizing up the Hoosier Mental Lexicon: Measuring the familiarity of 20,000 words. Research on Speech Perception Progress Report, 10, 357–376.

Patel AD (2008). Music, language, and the brain. Oxford University Press.

Raftery AE (1995). Bayesian model selection in social research. In Marsden PV (Ed.), Sociological methodology 1995 (pp. 111–196). Blackwell.

Ramus F, Nespor M, & Mehler J. (1999). Correlates of linguistic rhythm in the speech signal. Cognition, 73, 265–292. [PubMed: 10585517]

Redondo J, FragaI Padrón I, & Comesaña M. (2007). The Spanish adaptation of ANEW (Affective Norms for English Words). Behavior Research Methods, 39, 600–605. [PubMed: 17958173]

Regaiolli B, Rizzo A, Ottolini G, Miletto Petrazzini ME, Spiezio C, & Agrillo C. (2019). Motion illusions as environmental enrichment for zoo animals: A preliminary investigation on lions (Panthera leo). Frontiers in Psychology, 10, 2220. https://www.frontiersin.org/article/10.3389/fpsyg.2019.02220 [PubMed: 31636583]

Rogers JL, Howard KI, & Vessey JT (1993). Using significance tests to evaluate equivalence between two experimental groups. Psychological Bulletin, 113, 553–565. [PubMed: 8316613]

Rose EM, Mathew T, Coss DA, Lohr B, & Omland KE (2018). A new statistical method to test equivalence: An application in male and female eastern bluebird song. Animal Behaviour, 145, 77–85.

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. [PubMed: 30238294]

Schmidtke DS, Schröder T, Jacobs AM, & Conrad M. (2014). ANGST: Affective norms for German sentiment terms, derived from the affective norms for English words. Behavior Research Methods, 46, 1108–1118. [PubMed: 24415407]

Siakaluk P, Newcombe PI, Duffels B, Li E, Sidhu DM, Yap MJ, & Pexman PM (2016). Effects of emotional experience in lexical decision. Frontiers in Psychology, 7, 1157 10.3389/fpsyg.2016.01157 [PubMed: 27555827]

Siakaluk PD, Knol N, & Pexman PM (2014). Effects of emotional experience for abstract words in the Stroop task. Cognitive Science, 38, 1698–1717. 10.1111/cogs.12137 [PubMed: 24964820]

Siew CSQ (2019). Spreadr: An R package to simulate spreading activation in a network. Behavior Research Methods, 51, 910–929. [PubMed: 30788800]

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

Soares AP, Comesaña M, Pinheiro AP, Simões A, & Frade CS (2012). The adaptation of the Affective Norms for English words (ANEW) for European Portuguese. Behavior Research Methods, 44, 256–269. [PubMed: 21751068]

Solomon SG, & Lennie P. (2007). The machinery of colour vision. Nature Review Neuroscience, 8, 276–286. [PubMed: 17375040]

Tay RYL, & Ng BC (2019). Effects of affective priming through music on the use of emotion words. PLOS ONE, 14(4), Article e0214482.

Tierney A, Dick F, Deutsch D, & Sereno M. (2013). Speech versus song: Multiple pitch-sensitive areas revealed by a naturally occurring musical illusion. Cerebral Cortex, 23, 249–254. [PubMed: 22314043]

Tierney A, Patel AD, & Breen M. (2018). Repetition enhances the musicality of speech and tone stimuli to similar degrees. Music Perception, 35, 573–578.

Vanden Bosch der Nederlanden N, Hannon EE, & Snyder JS (2015). Everyday musical experience is sufficient to perceive the speech-to-song illusion. Journal of Experimental Psychology: General, 144, e43–e49. [PubMed: 25688906]

Vitevitch MS (2003). Change deafness: The inability to detect changes in a talker's voice. Journal of Experimental Psychology: Human Perception and Performance, 29, 333–342. [PubMed: 12760619]

Vitevitch MS, Chan KY, & Roodenrys S. (2012). Complex network structure influences processing in long-term and short-term memory. Journal of Memory & Language, 67, 30–44. [PubMed: 22745522]

Vitevitch MS, & Donoso A. (2011). Processing of indexical information requires time: Evidence from change deafness. Quarterly Journal of Experimental Psychology, 64, 1484–1493.

Vitevitch MS, Ercal G, & Adagarla B. (2011). Simulating retrieval from a highly clustered network: Implications for spoken word recognition. Frontiers in Language Sciences, 2, 369.

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

Vitevitch MS, & Luce PA (1998). When words compete: Levels of processing in spoken word perception. Psychological Science, 9, 325–329.

Vitevitch MS, & Luce PA (2004). A web-based interface to calculate phonotactic probability for words and nonwords in English. Behavior Research Methods, Instruments, and Computers, 36, 481–487.

Vitevitch MS, & Luce PA (2005). Increases in phonotactic probability facilitate spoken nonword repetition. Journal of Memory & Language, 52, 193–204.

Vitevitch MS, & Siew CSQ (2017). Estimating group size from human speech: Three's a conversation, but four's a crowd. Quarterly Journal of Experimental Psychology, 70, 62–74.

Vitevitch MS, & Sommers MS (2003). The facilitative influence of phonological similarity and neighborhood frequency in speech production in younger and older adults. Memory & Cognition, 31, 491–504. [PubMed: 12872866]

Walker E, & Nowacki AS (2010). Understanding equivalence and noninferiority testing. Journal of General Internal Medicine, 26, 192–196. [PubMed: 20857339]

Wilson M, & Cook PF (2016). Rhythmic entrainment: Why humans want to, fireflies can't help it, pet birds try, and sea lions have to be bribed. Psychonomic Bulletin & Review, 23, 1647–1659. [PubMed: 26920589]

Zhang S. (2011, August). Speech-to-song illusion in MC: Acoustic parameter vs. perception [Poster presentation]. Poster presented at the Biennial Meeting of the Society for Music Perception and Cognition, Rochester, NY, United States.

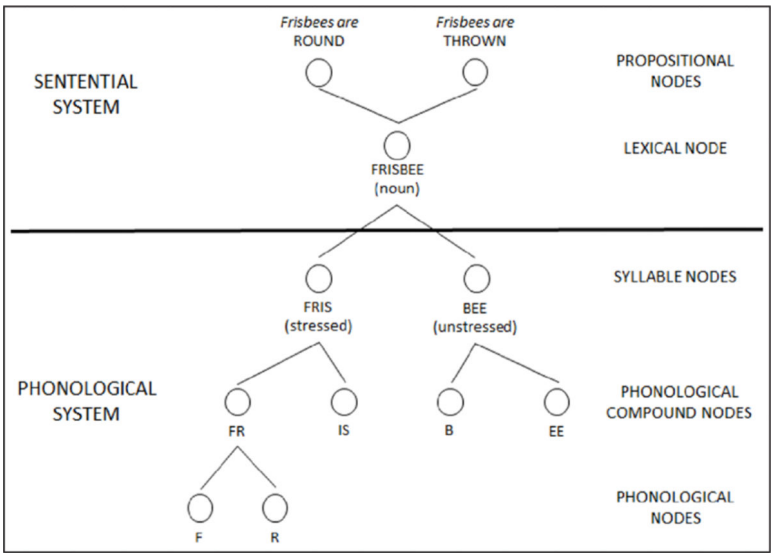

Figure 1.

The constituent phonemes and syllables, as well as the semantic information associated with the word *frisbee* as it might be represented in NST. Additional higher level and lower level nodes have been omitted to simplify the image.

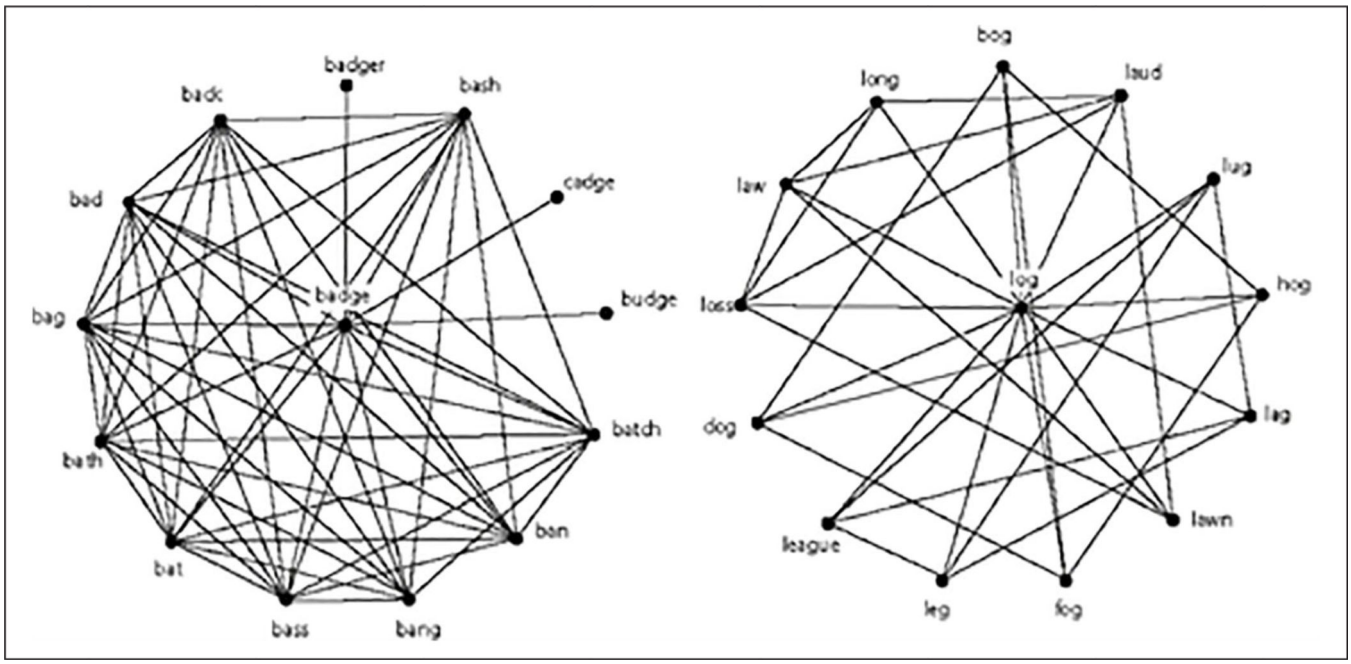

Figure 2.

The left panel represents a word with a higher clustering coefficient (*badge*), whereas the right panel represents a word with a lower clustering coefficient (*log*). Note that both words have the same number of phonological neighbours.