UNIVERSITEIT VAN AMSTERDAM

When Speech "behaves so strangely": The Influence of Prosody Type on the Speech-to-Song Illusion.

Martha Nobbe Smyth (12892173)

Universiteit van Amsterdam

Bachelor's Thesis Linguistics

Instructors: Dr. M. Sadakata and Dr. B.M. van 't Veer

26/06/2023

Abstract

It is usually taken as a given that humans can distinguish between when somebody is speaking or singing. A growing body of work investigating how the domains of music and language overlap, shows that the two not only share acoustic features such as pitch and rhythm but also have been shown to share neural mechanisms. This perceptual boundary between speech and song is elucidated by recent work into the so-called *speech-to-song illusion.* This illusion was presented in the preliminary study by Deutsch et. al. (2011), showing how repetition of a certain phrase would lead to it being perceived as song, as opposed to speech. Follow-up studies have shown that the illusion is mediated by pitch and rhythmic features of the phrase and is furthermore reduced in native speakers of a tone language compared to speakers of a non-tonal language. The current study investigates how speakers of a pitch-accent language experience this illusion, in order to observe whether the illusion is mediated by the prosodic system of one's language. In order to investigate this, a behavioural experiment was conducted with native speakers of Japanese and native speakers of Hiberno-English. The results of the study showed that contrary to the hypothesised reduced effect for Japanese speakers, both groups of speakers showed similar *speech-to-song* transformations. These findings are posited to potentially reflect a stronger mediating effect of language background than what is currently thought, proposing further cross-linguistic studies into this illusion.

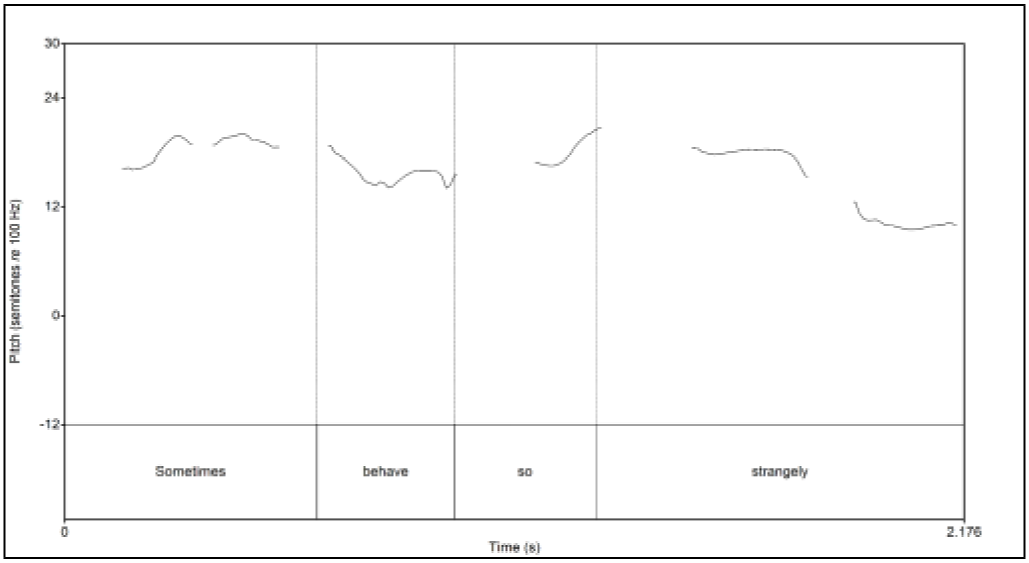
Table of Contents

1. Introduction

Why is it that we process speech as speech, and song as song? The domains of language and music both involve complex and meaningful sequences (Patel, 2010) with speech and song acting as prototypical examples of the two respectively. The similarities between speech and song have led to a substantial body of empirical evidence comparing and contrasting the two processes (Tierney et al., 2013; Zatorre & Gandour, 2007). Evidence suggests that the brain mechanisms involved in music and language processing interact, showing a bidirectional music-language transfer effect through experience in either domain (Bidelman et al., 2013).

A recent area of interest for the boundary between music and language processing is the so-called ‘speech-to-song’ (STS) illusion. The STS illusion refers to a dramatic shift in our perception of short speech fragments, which when repeated, may start to be perceived as more song-like than speech-like. The STS illusion was initially investigated via a behavioural experiment in a paper by Deutsch et al. (2011). In this paper, 54 participants with at least nine years of musical training took part in a two-part experiment. The first part of the experiment consisted of a listening experiment, where participants listened to the phrase “sometimes behave so strangely”, which was an extract of a female voice (Deutsch’s) taken from a CD recording. Participants then rated the presented stimulus on a Likert scale ranging from 1 to 5, marked from “exactly speech-like" to “exactly song-like”. The phrase was presented in three different conditions: unmanipulated, transposed for pitch, and with jumbled syllables. After the first phrase rating, the same phrase was presented again but was repeated ten times, with an interval of 2300 ms, which the participants again rated on the same Likert scale. Figure (1) shows the pitch contour of the unmanipulated stimulus used in the experiment.

FIGURE (1) | *Pitch contour of unmanipulated stimuli used in Deutsch et al. (2011)*



This process was conducted for all three conditions. The second part of the experiment included a production task, where participants imitated the same stimulus phrase before and after repetition. The main findings of this paper show that, in the listening condition, song ratings were low after initial hearing, but dramatically increased after repetition in the unmanipulated condition, as well as in the transposed pitch condition, the latter with a less dramatic increase. This finding reflects that pitch plays a role in the transformation of this illusion, as well as highlighting that this illusion requires certain acoustic features in order for the stimuli to transform. Furthermore, the results of the production task showed that imitation after repetition resulted in less variance in pitch (F0) in comparison to the pre-repetition phrase imitation, further supporting the hypothesis that repetition leads to an increase in song perception in certain phrases. In sum, this preliminary study into the STS illusion laid the foundations for what is now known about this perceptual illusion, providing compelling evidence that repetition induces a shift in our perception from speech to song with certain phrases.

This perceptual illusion has been explored in a number of empirical studies, including research into acoustic correlates of this illusion, showing that pitch is the most salient mediator of the transformation from speech to song (Tierney et al., 2013). Other research has also shown that there is a reduced STS illusion in native speakers of both Mandarin Chinese and Thai, both of which are tone languages, compared to native speakers of German and Italian, both non-tonal languages (Jaisin et al., 2016). What has not yet been investigated in the domain of the STS illusion, however, is how pitch-accent language speakers experience this illusion. Pitch-accent languages, (further discussed in Section 1.2) much like tone languages, make use of pitch to denote lexical meaning, however, unlike tone languages, they comprise a small number of contrasting tones (Yip, 2002). Investigating the STS illusion in pitch-accent languages would provide more insight into how the degree of tonality in a language influences the efficacy of the illusion, adding to the growing body of work elucidating the relationship between language experience and this illusion.

Moreover, research by Margulis et al. (2015) has shown that the language in which the stimuli are presented (henceforth stimulus language) plays a role in the strength of the speech-to-song transformation, with stimuli in a hard-to-pronounce language relative to one's native language showing stronger transformations than easy-to-pronounce languages (further discussed in Section 1.4). Given the findings of these empirical studies, the current study aims to address the gap in the field regarding pitch-accent, by observing the efficacy of the STS illusion in native speakers of the pitch-accented language, Japanese, compared to speakers of the stress-accent language speakers, Hiberno-English (the variety of English spoken in Ireland). Furthermore, the findings of Margulis et al. (2015) will be further investigated, by looking into whether native or non-native stimuli result in stronger STS transformations.

1.1 Acoustic Correlates of the Speech-to-Song Illusion

As outlined above, the STS illusion has been the topic of a myriad of empirical research into the boundary between music and language (Tierney et al., 2013; Jaisin et al., 2016; Margulis et al., 2015). Among these works are studies into the acoustic correlates of this STS illusion, focusing on the acoustic features of transforming phrases.

A study by Tierney et al. (2018) investigated these acoustic features by manipulating specific characteristics of speech to see which were causally related to perceiving a speech segment as sung. The characteristics observed included within-syllable pitch contour, measured by extracting the slope of each syllable's pitch contour, and melodic structure, and musical beat structure.

The study consisted of four experiments; one testing whether these characteristics were predictive of the speech-to-song transformation, and three testing the causal role of each of the observed characteristics. The experiment included 45 participants in total, who took part in all four experiments. The stimuli consisted of 48 short phrases taken from audiobooks, which were originally intended to be heard as speech, 24 which were chosen to elicit the illusion, 24 to act as a control. The first experiment presented these 48 stimuli to participants without altering the acoustic features of the stimuli. In each of the following experiments the stimuli were manipulated in order to isolate each of the proposed acoustic correlates. The listening and rating of the experiment echoed that of the preliminary study by Deutsch et al. (2011), in which participants rated the presented stimuli on a one to ten Likert scale, ranging from 'speech-like' to 'song-like', both pre and post repetition.

The results from Tierney et al's Experiment 1 showed a significant increase in song-likeness rating from the initial hearing of the stimuli to the repetition section, suggesting that repetition of certain stimuli increases the musicality of spoken phrases so that they will be perceived as more song-like than speech-like after they have been repeated. Furthermore, the increase in song-likeness ratings was correlated with beat variability and within-syllable pitch slope. The results of Experiments 2, 3 and 4 showed that only within-syllable pitch slope modulated the magnitude of the increase in song perception with repetition, with no significant effects of musical beat structure or melodic structure. This study thus elucidates the significant pitch and rhythmic features of speech relevant to the STS illusion. Namely, a stable/flat within-syllable pitch slope acting as the primary mediator of the transformation with regular beats mediating the transformation to a lesser degree.

The findings presented in the paper by Tierney et al. (2018) were alluded to in a paper by Falk et al. (2014), which also investigated the acoustic properties that mediate the efficacy of the STS illusion. Adding to the established finding that rhythmic features are determinants of the illusion, Falk et al. (2014) provide evidence that, while pitch properties are the most reliable cue for the transformation, regularly recurring temporal relations between accented and unaccented intervals in the speech segment also mediate the transformation.

In sum, there is a strong body of empirical evidence investigating the acoustic correlates of the STS illusion, which show that pitch and rhythmic features are modulators of the STS illusion. Specifically, a stable or flat within-syllable pitch contour acts as the most salient acoustic cue for the perceptual transformation, with regularly recurring temporal relations between accenting acting as a secondary modulator.

1.2 Language Tonality and the Speech-to-Song Illusion

The previous section highlighted that a number of empirical studies show that a stable pitch contour is the most salient feature in what conditions the STS illusion transformation. Pitch is defined as the perceptual correspondent of fundamental frequency (F0), which is the acoustic signal of vocal fold vibrations (Yip, 2002). Pitch is an essential component of both music and spoken language. It is one of the main dimensions of how music can be distinguished, conveying information about tonality, harmonics, phrase boundaries, rhythm, and meter (Pfordresher & Brown, 2009). Furthermore, it is one of the most salient aspects of an organised system of musical elements (Patel, 2010).

The notion of pitch in language systems falls under the category of prosody. Despite the fact that word prosody is a frequently investigated phenomenon in the field of linguistics, the classification of languages into stress-accent (SA), tone (T) and pitch-accent (PA) languages continues to pose as a challenge (Hyman, 2006). The definition of SA and T as presented by Hyman (2006) are as follows:

> **(1) SA**: A language with stress accent is one in which there is an indication of word-level metrical structure meeting the following two criteria:
>
> > *Obligatoriness*: every lexical word has at least one syllable marked for the highest degree of metrical prominence.
> >
> > *Culminativity*: every lexical word has at most one syllable marked for the highest degree of metrical prominence.

By this definition, languages with word-level metrical structure, such as English, are considered stress-accent languages. SA systems are realised not by one phonetic feature but can be potentially realised by multiple, such as duration, vowel quality and pitch (Gussenhoven & Gussenhoven, 2004; Hyman, 2006). This adds to the difficulty in classifying languages strictly into the three aforementioned categories. SA systems are, however, classified strictly as assigning prominence to syllables, specifically minimally an obligatory head syllable per word, ruling out any languages which assign prosody to morae, such as Japanese, Safwa and Kinga (Hyman, 2006).

(2) T: A language with tone is one in which an indication of pitch enters into the lexical realisation of at least some morphemes

A tone language is thus defined as a language with word-level pitch features, such as Mandarin Chinese or Thai, which, unlike SA languages, are realised by only one phonetic feature: pitch, which is used to express lexical contrasts. Regarding the definition of PA language systems, a number of definitions have been put forward, such as a very reduced tone system (Hyman, 2006), and a language in which lexical tone does exist. However, unlike tone languages, they comprise a small number of contrasting tones (generally one or two) (Yip, 2002). While presenting a number of possible descriptions for the PA system, the paper by Hymann (2016) concludes that no independent definition can be put forward due to the abstractness of accent, as well as the fact that PA systems vary with regard to their inclusion of stress and tone.

A more recent approach to defining these three categories, SA, T and PA, has been to consider the relationship between the three as a multi-dimensional continuum, accounting for tonal density, i.e. the percentage of prosodic units which require a tonal feature, and criterion inclusion (such as obligatoriness and culminativity) (Hyman, 2016). While not accounting for metrical structure, the continuum presented in Figure (2) gives a shallow but clear image of how these language systems would be ranked on a continuum when observing tonal density (Hyman, 2016).

FIGURE (2) | *Prosodic category continuum based on language tonal density*

English-------W. Basque-------Tokyo Japanese-------Luganda-------Mandarin Chinese

Figure (2) shows a continuum ranging from no lexical tone for the SA system on the left to the T systems on the right. Under the definition by Hymann (2016), the languages W. Basque, Tokyo Japanese, and Luganda are classed under the PA system, with the Japanese language being arguably the most frequently investigated language in the pitch-accent category. The lexical tones in a TA system usually belong to specific syllables or morae and may be either sparsely distributed or absent in some words, with the proportion of words that contain pitch-accent in Japanese at just over 50 per cent (Gussenhoven & Gussenhoven, 2004). An illustration of how pitch is used in Japanese is provided in Table (1) below, highlighting the restricted tone use in Japanese, consisting only of high (H) and low (L) tones (Tsujimura, 1999).

Table (1) | ***Example of tone use in Japanese (adapted from, Tsujimura, 1999)***

Unaccented	***Initial-accented***	***Second-accented***
hasi(-ga)[1]	ha`si(-ga)	hasi`(-ga)
LH H	HL L	LH L
“Handle”	“Chopstick”	“Bridge”

Considering the substantive amount of work which classifies Japanese under this pitch-accent system, native Japanese speakers will be recruited to represent speakers of a pitch-accent language for the purpose of this study. Japanese comprises a number of dialects, most of which use lexical pitch, however, there are a number of dialects which are deemed accentless, including Southern Miyagi, Southern Yamagata and the Fukushima dialect (Sato et al., 2013). This difference in accenting across Japanese dialects will be accounted for in this study by asking participants which dialects they speak in a post-experiment questionnaire.

1.2.1 Influence of Tone Language on Speech-to-Song Transformation

This distinction between the three proposed systems, SA, T and PA, presents an interesting case for the investigation into the STS illusion, as it gives an insight into the distinct profiles of lexical pitch processing across the three language types. Furthermore, the question as to the extent to which this distinction in lexical processing can mediate the effect of the perceptual transformation from speech to song is raised. This relationship has previously been investigated by Jaisin et al. (2016), who conducted a study on whether natively speaking a tone language influences the propensity for the STS illusion to transform, compared to speakers of a

[1] *ga* = Nominative morpheme in Japanese (Tsujimura, 1999)

non-tonal language. The authors conducted a behavioural experiment, including native speakers of Mandarin Chinese and Thai, to test the effect on native tone language speakers, and native speakers of German and Italian, as a control non-tonal language speaking group. While the STS illusion had previously been tested with native German speakers, showing a significant transformation for German stimuli (Falk et al., 2014), the illusion had not previously been tested on speakers of Italian, Mandarin Chinese or Thai.

All participants were L2 speakers of English, which was included in the stimuli set, in order to provide a reference for interpreting any potential language-specific effects. Participants were presented with stimuli that had previously been shown to generate the STS illusion, presented in all five languages: Mandarin Chinese, Thai, German, Italian and English. All languages were presented to the participants in order to observe the previously established finding by Margulis et al. (2015), showing higher transformation scores for stimuli in hard-to-pronounce languages compared to ones native language.

The main findings of the paper showed that the STS illusion is reduced in native speakers of a tone-language compared to native speakers of a non-tonal language, with the overall strength of the speech-to-song effect across all language stimuli being relatively weak for tone-language speakers. The author posits that these results could indicate that tonality of native language may determine the extent to which prosodic pitch features are perceived as conforming to musical or lexical melody. To conclude, the authors state that their findings are to be regarded as evidence that the STS illusion is reduced in native speakers of tone languages, proposing that further work including larger cohorts of tone language sampling would be necessary in order to establish the reliability of the findings presented in their paper.

Considering the evidence provided by Jaisin et al. (2016), the current paper aims to investigate the gap in the field of the relationship between languages with lexical tone and the STS illusion, by investigating whether the illusion is also reduced in speakers of a pitch-accent language. Empirical evidence from neuroimaging studies comparing pitch processing in tone language and pitch-accent language speakers indicates that the two show similar pitch processing, (Sato et al., 2007). Considering this, as well as the findings by Jaisin et al. (2016) a reduced STS illusion effect is hypothesized for native speakers of a pitch-accent language compared to speakers of a stress accent language, such as English.

1.3 Neural Underpinnings of Pitch Processing in Speech and Song

Given the aforementioned overlapping acoustic features of both music and language, the two domains have previously been explored with regard to high-level brain organisation via electroencephalogram (EEG) and functional magnetic resonance imaging (fMRI) studies, for example. While the vast majority of studies investigating the STS illusion consist of behavioural studies, the neural underpinnings of the perceptual transformation were explored via an fMRI study by Tierney et al. (2013). In this paper, 14 native English-speaking participants took part in a brain imaging experiment, where they were exposed to three blocks of differing stimuli: the speech block, in which previously attested non-transforming phrases were presented, the song block, in which previously attested transforming phrases were presented, and finally, the silence block, in which no stimulus was played, as a control . The stimuli were played through CONFON headphones, which participants wore not only to listen to the stimuli, but also to dampen the noise of the fMRI scanner, controlling for distracting acoustic interference. Both speech and song blocks consisted of a 16-second stimulus repetition listening task, with a 500ms interstimulus interval between each repetition. The blocks were presented in pseudorandom order. During this listening task, participants were asked to mentally note whether the phrases sounded more speech-like, or song-like, while their brain region responses were being recorded via fMRI.

The main findings of the study show that, while there were no regions more highly responsive to speech than to song, there were multiple brain regions that were more responsive to song than to speech, consisting of areas that can broadly be divided into areas responsible for pitch processing, and areas involved in vocalization and auditory-motor integration. Furthermore, increased responses in song perception did not lead to an increase in response to the primary auditory cortex. The authors indicate that these results reflect the processing differences between speech and song, with increased demands on neural resources for song perception than for speech perception. Moreover, these results support the claim made in the behavioural study by Deutsch et al. (2011), which puts forward the idea that the STS illusion reflects a shift in our perception from speech to song after repetition.

What the study by Tierney et al. (2013) also highlights is the processing differences of pitch in music and language, showing differing brain activities for pitch processing in the two domains, conditioned by specific language experiences. This language-experience influence on divergent pitch processing activities was explored by Zatorre & Gandour (2007) where the authors give a state-of-the-art account of pitch processing differences in speakers of tonal and non-tonal languages. While the authors acknowledge that a full understanding of pitch processing across languages, as well as pitch processing in music, is still not fully available, a strong body of evidence points to a divergence in hemispheric processing, determined by one's language experience of tonal or non-tonal languages. Specifically, the authors posit that in the case where pitch patterns are phonologically significant, such as in tone languages, the pitch processing is left-hemisphere lateralised; when it is not phonologically significant, such as in non-tonal languages, pitch processing in right-hemisphere lateralised. Although this paper does not directly address the processing of pitch in pitch-accent languages, it points towards left-lateralisation, similar to tone languages due to pitch being phonologically significant in both.

Moreover, empirical research into the processing of lexical pitch in both Mandarin Chinese and Standard Japanese, suggests that speakers of the two share similar neural underpinnings of prosodic processing. A paper by Sato et al. (2007) investigates the neural correlates of lexical pitch-accent processing in native speakers of Japanese, hypothesising that if Japanese lexical pitch is processed like tones in Chinese and Thai, as elucidated by Zatore & Gandour (2007), native Japanese speakers will show greater activation of the language-related left frontal and tempo-parietal regions. 20 native speakers of the Tokyo dialect of Japanese took part in a near-infrared spectroscopy recording experiment, in which they were presented with sets of Japanese minimal pairs, that differed in pitch contour (HL vs LH). The stimuli consisted of a four-condition block design paradigm, including (1) accent condition, (2) phoneme condition, (3) variable words condition and (4) pure tone condition. Participants' hemodynamic responses to the stimuli were recorded whilst the block design paradigm was presented. The results of this study show that native Japanese speakers show higher activation in the left tempo-parietal region, as well as the left frontal region during the perception of pitch pattern changes associated with lexical items. The results presented in this study align with empirical work on the processing of lexical tone in tone languages (Zatore & Gandour, 2007), highlighting how both pitch-accent language prosody and tone language prosody are processed in the left hemisphere, or language-specific domain.

This finding is of interest as it shows, despite the divergence in how pitch is used in tone languages such as Mandarin Chinese and Thai, to pitch-accent languages such as Japanese, that there are apparent similarities in the processing of pitch. Investigating the effectiveness of the STS illusion on native Japanese speakers would thus give insight as to whether the degree of tonal use in a language influences this illusion or not.

1.4 Influence of Demographic Variables on the Speech-to-Song Illusion

Considering the fact that the STS illusion highlights the perceptual boundary between music and language, a number of studies have investigated the influence of external factors on the illusion, such as the influence of stimulus language, and musical experience/training, (Margulis et al., 2015; Tierney et al., 2021).

A paper by Margulis et al. (2015) looked into the effect of stimulus language on the propensity for the STS illusion to be elicited, by observing whether stimuli in harder-to-pronounce languages relative to one's native language resulted in lower transformation scores than stimuli in a native language. The motivation for investigating this influence was based on the empirical finding that transforming stimuli result in increased demands on neural resources compared to non-transforming stimuli (Tierney et al., 2013), which the authors propose could reflect a participatory stance of the listener, where they begin to sing the tune in their head once it starts to transform. Based on this, the authors hypothesise a stronger transformation for stimuli in one's native language.

A behavioural experiment was carried out in order to test this hypothesis, including 24 native English-speaking participants. The stimuli were created based on one English phrase, which was then translated into Catalan, Portuguese, French, Croatian, Hindi and Irish, resulting in stimuli in seven different languages (including English as the control stimulus). While the authors did not include the participant's various language backgrounds aside from their L1 (English, with multiple fluent speakers of a different L2), they conducted a post-experiment questionnaire in order to account for how hard to pronounce the participants deemed each of the stimuli. The results showed that Catalan and Portuguese were considered easy to pronounce by the English-speaking participants, French and Croatian were considered relatively harder to pronounce, and Hindi and Irish were considered hard to pronounce. In order to test for the *speech-to-song* transformation, the study utilised a similar set-up to the one presented in the initial study by Deutsch et al. (2011), with a Likert scale from 1, "*sounds exactly like speech",* to 5, "*sounds exactly like song".* The listening and pre and post-repetition rating procedure also duplicated the preliminary STS illusion experiment by Deutsch et al. (2011). The results of the behavioural experiment showed that contrary to the proposed hypothesis, the harder-to-pronounce languages were more susceptible to the STS transformation, showing that the more difficult participants rated the pronunciation of the language, the higher transformation scores they showed. What the authors propose regarding the outcome of the experiment, is that languages that are native, or rated relatively easy to pronounce are captured most successfully by the speech processing circuitry, making them more resistant to being processed by other perceptual mechanisms. The results of the paper by Margulis et al. (2015) were further replicated in a later study by Castro et al. (2018), which provides evidence for a stronger STS transformation with repeated speech in an unfamiliar language.

Additionally, studies investigating whether musical experience plays a role in how successfully the STS illusion transforms show that musicians and non-musicians show no significant difference in their STS transformations (Vanden Bosch der Nederlanden et al., 2015; Tierney et al., 2021). While the paper by Tierney et al. (2021), provides evidence in favour of a universally transforming effect across both musicians and non-musicians, the authors did show that the ability to direct attention to pitch and beat perception during the illusion was a predictor of how strongly the illusion elicits a transformation. What is thus concluded from this paper, is that while musical training or experience does not mediate the perceptual transformation, musical aptitude can be considered a key factor in determining how strongly the transformation is experienced. Contrary to the findings of Vanden Bosch der Nederlanden et al. (2015) and Tierney et al. (2021), however, a paper by Groenveld et al. (2020) showed that musical sophistication, determined via a Gold-MSI self-report inventory (Müllensiefen et al., 2014), resulted in a reduced *speech-to-song* transformation. Unlike the binary splitting of musicians and non-musicians used in previous investigations into demographic variables, this alternative way of modelling the effect of musical background for the STS illusion enables the observation of particular aspects of musical background which influence the transformation effect.

In sum, the current body of work into the influence of demographic variables on the STS illusion shows that stimulus language influences the strength of the perceptual transformation, with languages that are harder-to-pronounce relative to one's native language showing stronger transformations. Furthermore, while there is evidence showing similar experiences of the STS illusion across musicians and non-musicians, there is also evidence that musical sophistication does play a role in how successful the illusion is. Considering the findings of these empirical studies, the current paper aims to further investigate the effect of stimulus language on how effective the illusion transforms, by including monolingual Japanese and Hiberno-English speakers and presenting stimuli in Japanese and English to both groups. Moreover, a subset of questions from the Gold-MSI self-report inventory will be included in a post-experiment questionnaire, in line with the work of Groenveld et al. (2020).

1.5 Current Study

Considering the gap in the field with regards to how natively speaking a pitch-accent would influence the efficacy of the STS illusion, the primary aim of the current study is to investigate the relative strength of the illusion in native Japanese speakers, by asking the research question: *whether the speech-to-song illusion is reduced in native speakers of a pitch-accent language compared to speakers of a stress-accent language.* In line with the findings of Jaisin et al. (2016) and Sato et al. (2007), the speech-to-song illusion is predicted to be reduced in native speakers of a pitch-accent language compared to speakers of a stress-accent language. Moreover, in order to further investigate the findings of Margulis et al. (2015), the following research question is proposed: *whether stimuli in a non-native language result in higher STS transformation scores than stimuli in a native language*. Higher transformation scores for stimuli in a non-native language compared to stimuli in a native language is hypothesised.

The proposed research questions will be explored via a listening-and-rating behavioural study replicating the set-up of established studies on the STS illusion (Deutsch et al., 2011; Tierney et al., 2013; Margulis et al., 2015). Participants will consist of native Japanese speakers, for the PA language group, and Hiberno-English for the SA language group. Hiberno-English is the variety of English spoken in Ireland. While there has been little explicit study done on the prosody of this vernacular, there is evidence that there is not much variation between stress-assignment between Hiberno-English, and other vernacular of English in the British Isles (Grabe et al., 2005), eluding to a similar SA system across these variations of English.

The goal in undertaking this study is to provide further insight into how language background and stimulus language mediate the efficacy of the STS illusion, with the broader goal of narrowing the understanding of the overlap in perception between language and music.

2. Methodology

This study was approved by the ethics committees of the Faculty of Humanities of the University of Amsterdam, filed and approved under the alphanumeric code: FGW-893_2023.

2.1 Participants

Thirty-nine adult speakers between the ages of 18 and 29 years old (mean: 24.6) participated in the experiment: 19 native speakers of Japanese, and 20 native speakers of Hiberno-English acting as a control group. Only monolingual speakers were included as to avoid the potential influence of L2 experience on the results of the experiment. Data from three participants were disregarded as they were outside of the age range specified, as well as data from one native Japanese speaker as they resided in the US, where exposure to English could not be controlled. In total, data from 35 participants was included in the final analysis: 16 native monolingual speakers of Japanese, and 19 native monolingual speakers of Hiberno-English.

All participants reported normal or corrected to normal hearing, with no history of neurological impairment, language disorder, or learning disability. Participants were recruited through word-of-mouth promotion, as well as promotion via personal social media platforms. Informed consent (Appendix 6.1) was obtained from all participants prior to the experiment.

2.2 Stimuli

The phrases for the Japanese stimuli set were obtained via personal communication with a Japanese collaborator (Tanaka, personal communication, April 2023). The stimuli consisted of 12 Japanese utterances, all spoken by a native female Standard Tokyo Japanese speaker, whose dialect includes pitch-accent. The stimuli were further cut into shorter speech segments via Praat software (Boersma & Weenick, 2023) in order to match the Tierney stimuli set for duration as closely as possible. The cut segments were further edited to comply with natural-sounding speech breaks, in accordance with advice from a native Japanese speaker.

In order to test whether the Japanese stimuli had the potential to elicit the STS illusion, a pre-test experiment was run, including seven native English speakers. All of the previously mentioned edited Japanese stimuli were presented in the same initial listening, repetition and Likert-scale response sections as will be described in Section 2.3 below for the main experiment. The results from the native English speaking participants in the pre-test showed that 15 of the tested stimuli successfully elicited the illusion and were thus utilized for the purpose of the main research experiment. The Japanese stimuli had an average duration of 1.39 s and an average pitch of 270.5 Hz.

The phrases for the English stimuli set were derived from the study by Tierney et al. (2013) in which 24 spoken phrases were shown to have a strong speech-to-song effect after repetition. These phrases were obtained from audiobooks and taken from passages intended to be heard as speech, recorded by multiple male and female voices. In order to match the number of stimuli in both language sets, the 15 most successfully transforming[2] stimuli from the Tierney et al. (2013) set were used. These 15 stimuli had an average duration of 1.18 s, and an average pitch of 146.58 Hz. Figures (3) and (4) illustrate an example of the pitch contour of the Japanese and English stimuli. Further stimuli details can be found in Appendix 6.3.

[2] Successfully transforming stimuli from the study by Tierney et al. (2013) were ranked from (1) to (24), with (1) showing the highest S2S transformation scores. Stimuli ranked (1) to (15) were included for the purpose of this study.

FIGURE (3) | *Pitch Contour Visualisations of Japanese stimulus JP0003*

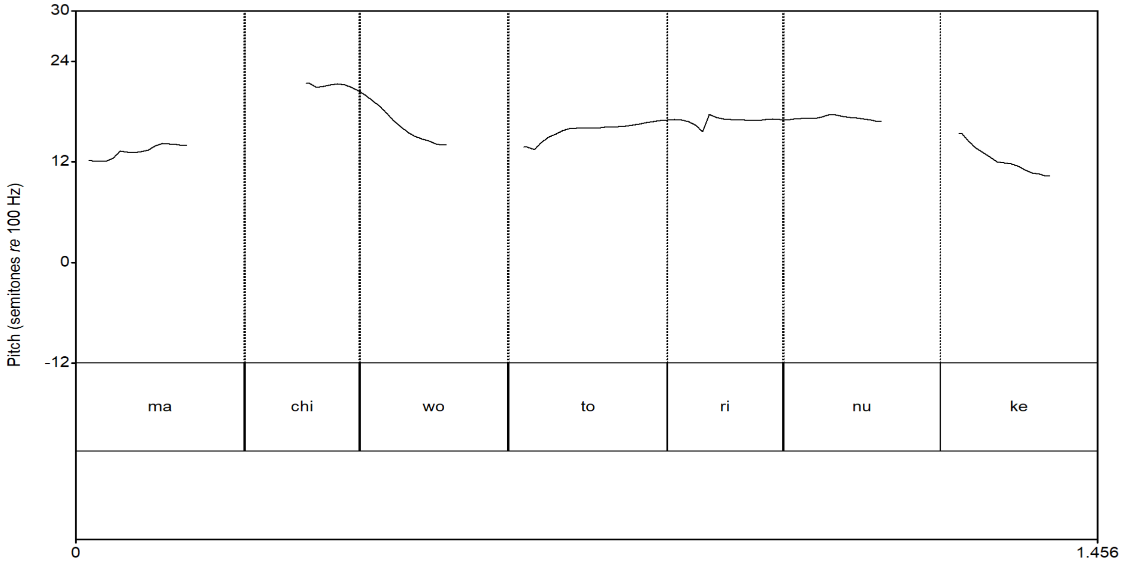
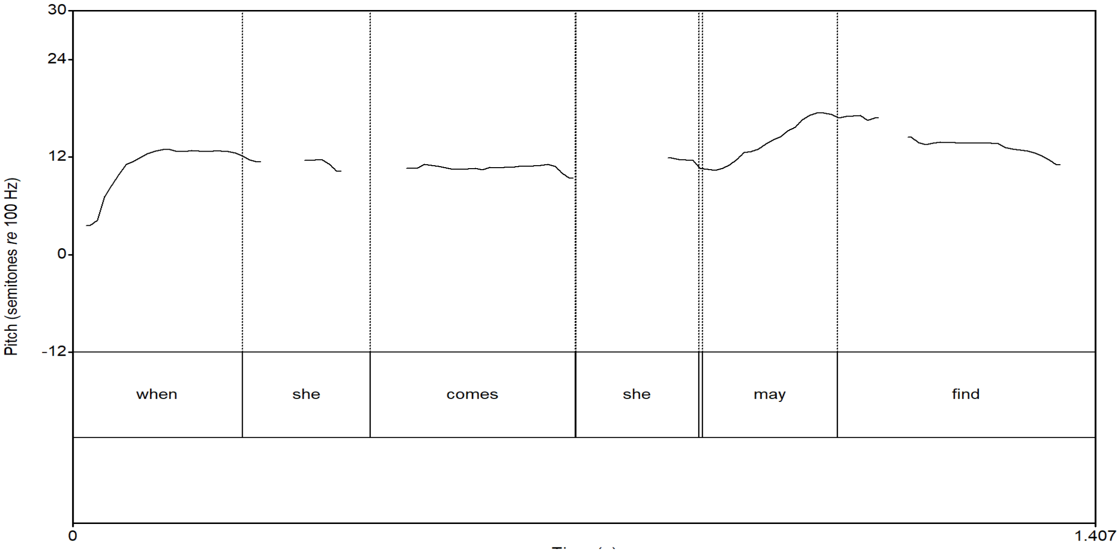


FIGURE (4) | *Pitch Contour Visualisations of English stimulus AT0006*



All of the stimuli from both the English and Japanese sets were normalized for volume and edited via Praat to remove any abrupt artefacts of cutting, in order to keep the stimuli as naturally speech-like as possible. The full list of both stimuli sets can be found in the appendix.

TABLE (2) | *Stimuli duration and pitch mean & standard deviation*

	Duration: Mean (s)	***Duration: SD (s)***	***Pitch: Mean (Hz)***	***Pitch: Mean SD (Hz)***
Japanese Stimuli	1.39	0.24	270.5	65.56
English Stimuli	1.180	0.28	146.58	22.99

2.3 Procedure

The experiment was administered to participants online, via Experiment Designer software (Vet, 2023) with instructions to carry out the experiment with headphones on, in a quiet environment. The online experiment consisted of two sections. The first, a listening experiment, consisted of 60 stimuli, the finalised stimuli in Japanese and English, as discussed in the preceding section, which were presented twice; once in the initial hearing, and again in the repetition section, played in sequence. The second part of the experiment consisted of a questionnaire, asking about the participant's language background and musical experience.

For each of the stimuli, participants were presented with (1) an initial hearing, followed by (2) a repetition section. During the repetition section, each speech segment stimulus was repeated 8 times in sequence, with a 400ms pause between each of the repetitions in order for the conditions to allow for the STS illusion to be elicited most successfully, in line with Falk et al. (2014). After both sections (1) and (2), participants were presented with a Likert-scale ranging from 1 to 7, with 1 corresponding to "speech-like", and 7 corresponding to "song-like", and were asked to rate the stimulus both pre and post-repetition according to the Likert-scale. Stimuli were presented in a randomized blocked order in an ABBA or BAAB format (Japanese, English, English, Japanese, or English, Japanese, Japanese, English) across the participant groups.

Once the participants had listened to and rated all 60 stimuli, they were asked to take part in a short questionnaire. The participants were asked for their age and gender, followed by a number of questions pertaining to their language background, including which dialect of English or Japanese they spoke. The final part of the questionnaire consisted of a sub-scale of questions extracted from the Gold-MSI self-report inventory, as mentioned in Section 1.3, pertaining to participants' musical sophistication and training. Both English version (Müllensiefen et al., 2014), as well as the validated Japanese version (Sadakata et al., 2022), were utilised in this study.

Finally, participants were presented with an open question, pertaining to their experience of the illusion in order to be able to elucidate on and discuss unexpected results if the hypotheses were not borne out. The experiment process took between 15-20 minutes in total. Both versions of the questionnaire, English and Japanese, can be found in the appendices (Section 6.4).

2.4 Data Analysis

In accordance with the literature, a significant effect of *Native Language*, with the native English-speaking group showing higher STS transformation scores than the native Japanese-speaking group, is expected (Jaisin et al., 2016). Furthermore, there is an expected significant interaction between the two independent variables; *Native Language* and *Stimulus Language*, reflecting higher transformation scores for stimuli presented in the non-native language compared to stimuli in the native language (Margulis et al., 2015).

As mentioned, the participant's responses to the experiment were measured by responding to a Likert scale ranging from 1 to 7. These ratings were recorded both pre- and post-repetition. For the purpose of this experiment, the STS illusion was defined operationally as the difference in song-likeness rating post-repetition compared to the initial pre-repetition rating (henceforth *transformation score*). The effect of the illusion will be tested for significance for each *Native Language* x *Stimulus Language* (Native English x English Stimuli, Native English x Japanese Stimuli, Native Japanese x English Stimuli, and Native Japanese x Japanese Stimuli), via a paired t-test on the pre and post-repetition ratings (see Table (3)).

In order to test for the abovementioned hypothesised effects, a two-way mixed ANOVA was employed, with *Native Language* (Japanese or English) as an independent, binary, between-group variable, *Stimulus Language* (Japanese or English) as an independent, binary, within-group variable and transformation score, as the continuous dependent variable.

TABLE (3) | *Analysis Scheme: Transformation scores for Native Language x Stimulus Language*

	Native English Speakers	*Native Japanese Speakers*
English Stimuli	*Native English x English Stimuli*	*Native Japanese x English Stimuli*
Japanese Stimuli	*Native English x Japanese Stimuli*	*Native Japanese x Japanese Stimuli*

Regarding the two-way mixed ANOVA analysis results, in the case that the English native-speaking group show a correlation with high song-likeness rating, and the native Japanese-speaking group does not, the hypothesis, that the STS illusion is reduced in speakers of a pitch-accented language compared to speakers of a non-pitch-accented language, is borne out. In the case that there is no significant cross-group difference in song-likeness rating, the null hypothesis, that the STS illusion is not reduced in speakers of a pitch-accented language compared to speakers of a non-pitch-accented language, is borne out.

In the case that there is a significant interaction between *Native Language* and *Stimulus Language*, the hypothesis that stimuli in a non-native language will show higher speech-to-song transformation scores than stimuli in a native language is borne out. If there is no significant interaction between *Native Language* and *Stimulus Language*, the null hypothesis, that unfamiliar language stimuli do not lead to a stronger STS illusion than stimuli presented in an unfamiliar language, is borne out.

3. Results

The data for each of the participants were collected in an Excel file, sorted by the first independent variable: *Native Language,* the second independent variable: *Stimulus Language*, and finally by the dependent variable; the speech-to-song *transformation* scores. All statistical tests were conducted using the programme *RStudio* (Posit team, 2023).

3.1 Transformation Effect

For the purpose of this study, the STS illusion effect was defined operationally as any difference in post-repetition scores compared to the pre-repetition scores, which was obtained by subtracting the post-repetition Likert-scale score from the initial/pre-repetition Likert-scale score. The data from both groups showed to be normally distributed, which was tested via a Kolmogorov-Smirnov Test (p-value 0.3187). In order to confirm a transformation effect across all groups, a paired t-test was conducted on the pre-repetition and post-repetition scores of the four groups:

(1) Native English speakers rating English stimuli,
(2) Native English speakers rating Japanese stimuli,
(3) Native Japanese speakers rating English stimuli, and finally
(4) Native Japanese speakers rating Japanese stimuli.

The results of the four paired t-tests can be seen in Table (4), which shows the average transformation scores of each of the groups, as well as their significance. Given the significance of the four groups, all being below 0.01, it can be concluded that all groups showed a significant transformation score.

TABLE (4) | *Transformation scores across both Native Language & Stimulus Language*

Native Language	***Stimulus Language***	***Mean***	***SD***	***t-statistic***	***DF***	***p-value***
English	**English**	1.189	1.342	-15.557	284	<0.01
	Japanese	0653	1.139	-9.671	284	<0.01
Japanese	**English**	1.9	1.923	-16.267	239	<.001
	Japanese	1.204	1.655	-11.867	239	<.001

3.2 Results of the Two-way Mixed ANOVA

In order to examine the main research question: *Whether the speech-to-song illusion is reduced in native speakers of a pitch-accent language compared to speakers of a stress-accent language* and the sub-research question: *whether stimuli in a non-native language resulted in higher speech-to-song transformation percentages than stimuli in a native language,* a two-way mixed ANOVA was conducted, with the variables as explicated in Section 2.4.

3.2.1 Effect of Native Language

In line with the hypothesis for the main research question; that the STS illusion is reduced in speakers of a pitch-accented language compared to speakers of a stress-accent language, a main effect of Native Language was predicted, with native speakers of Japanese showing lower transformation scores compared to native English speakers. The results show that there was no significant effect of Native Language (F(1) = 4.086, p = 0.0515) meaning that there is not sufficient evidence to reject the null hypothesis. Namely, the STS illusion is not reduced in speakers of a pitch-accented language compared to speakers of a non-pitch-accented language. Speech-to-song transformation effects are summarized for Native Language across the combined stimulus types in Table (5).

TABLE (5) | *Results of two-way mixed ANOVA: Native Language x transformation scores*

	Df	***Sum sq***	***Mean sq***	***f-value***	***p-value***
Native Language	1	103.8	103.8	4.086	0.052

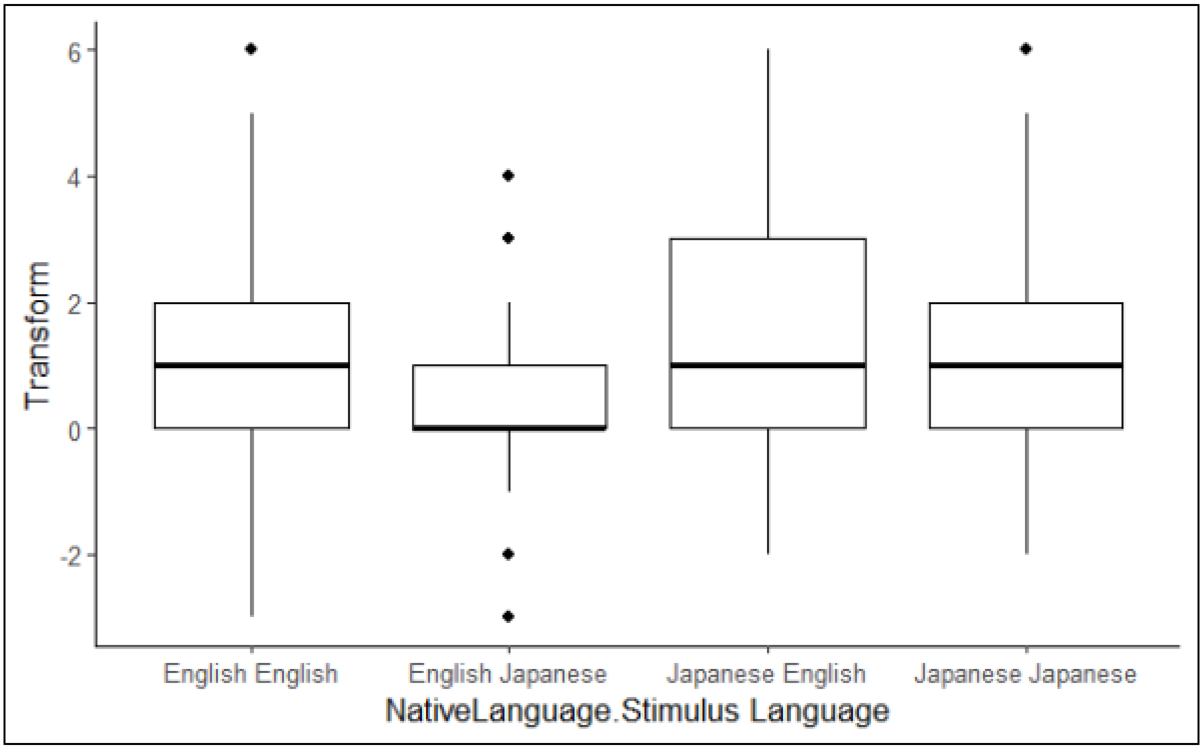
3.2.2 Effect of Stimulus Language

Regarding the sub-research question, the hypothesis, that stimuli in a non-native language would show higher transformation scores than stimuli in a native language, was put forward. Higher transformation for Japanese stimuli than the English stimuli for the native English-speaking group, and higher transformation for the English stimuli than the Japanese stimuli for the native Japanese-speaking group was thus expected. The hypothesis for the sub-research question would be borne out with no significant effect of *Stimulus Language,* and a significant interaction between *Native Language* and *Stimulus Language.* The results of the two-way mixed ANOVA showed that there was a significant main effect of *Stimulus Language*, showing that English stimuli (*M* =1.514, *SD* = 1.669) showed significantly higher transformation scores than the Japanese Stimuli (*M* = 0.905, *SD* = 1.424, *F*(1) = 29.374, *p* = <.001) across both groups of speakers. STS transformation effects are summarized for *Stimulus Language* across the combined participant groups in Table (6). Figure (5) shows the mean STS transformation scores across both independent variables, *Native Language* and *Stimulus Language* in a box plot The data points located outside the whiskers of the b plot represent the outliers.

TABLE (6) | *Results of two-way mixed ANOVA: Stimulus Language x transformation scores*

	Df	***Sum sq***	***Mean sq***	***f-value***	***p-value***
Stimulus Language	1	97.52	97.52	29.374	<.001

FIGURE (5) | *Mean speech-to-song transformation ratings for each Native Language and Stimulus Language.*



3.2.3 Interaction between Native Language & Stimulus Language

As previously mentioned, an interaction between Native Language and Stimulus Language was hypothesised. The results of the two-way mixed ANOVA analysis showed no significant interaction between the two independent variables, meaning that the initial hypothesis for the sub-research question hypothesis was not supported by the data. Table (7) shows the interaction between Native Language and Stimulus Language across all transformation scores.

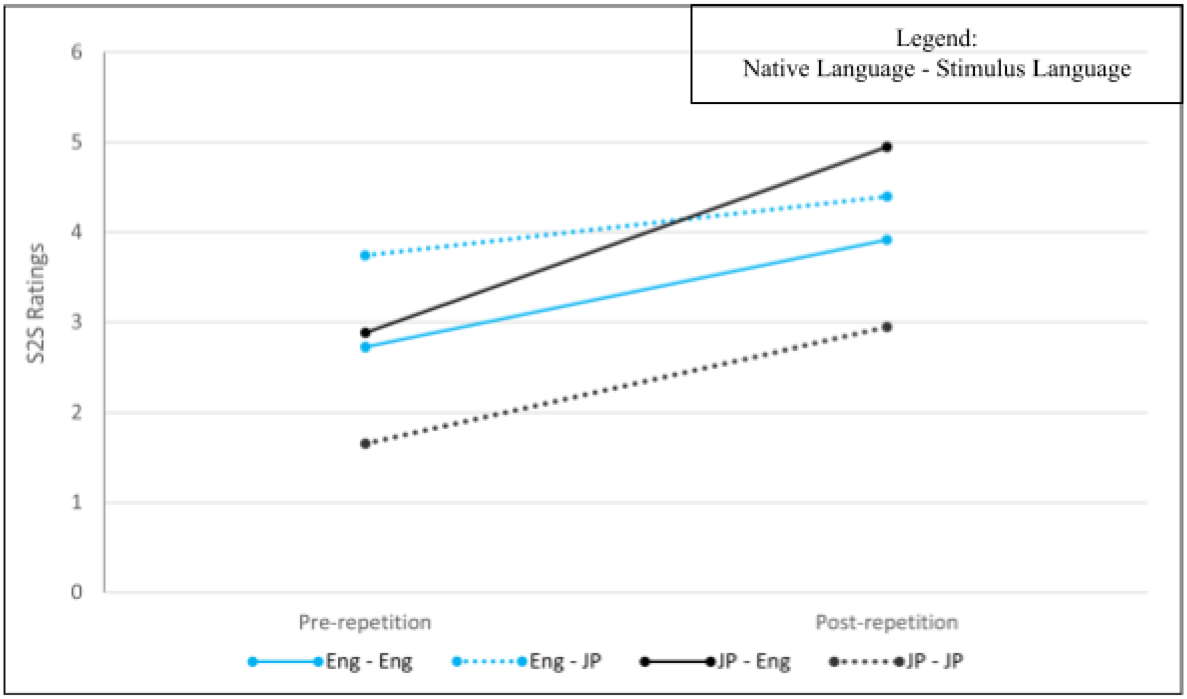
TABLE (7) | *Results of two-way mixed ANOVA: Native Language x Stimulus Language*

	Df	***Sum sq***	***Mean sq***	***f-value***	***p-value***
Native Language * Stimulus Language	1	1.65	1.65	0496	0.486

3.3 Summary of Main Findings

In sum, the results revealed that the two groups, native Japanese speakers, and native Hiberno-English speaking groups did not significantly differ from each other in speech-to-song transformation scores across all stimuli. Moreover, there was a significant main effect of Stimulus Language, with the English stimuli exhibiting significantly higher transformation scores than the Japanese stimuli across both native speakers of Hiberno-English and Japanese. Finally, there was no significant interaction between the two independent variables Native Language and Stimulus Language. The mean ratings for the groups (1)-(4) as described in Section 3, are presented in Figure (6), showing the trajectory from pre-repetition score to post-repetition score. From this graph, we can see the native English-speaking group rated the Japanese stimuli with a high “song-like” score for the initial hearing. The ratings across both groups for the English stimuli show similar initial hearing “song-like” scores, with the native Japanese speakers showing a stronger transformation than the native English-speaking group.

FIGURE (6) | *Mean speech-to-song transformation ratings for pre-repetition and post-repetition ratings. Blue coloured lines represent native English speakers, black coloured lines represent native Japanese speakers, solid lines represent English stimuli, dotted lines represent Japanese stimuli.*



4. Discussion

In this study, the research questions; *whether the speech-to-song illusion is reduced in native speakers of a pitch-accent language compared to speakers of a stress-accent language* and *whether stimuli in a non-native language resulted in higher speech-to-song transformation scores than stimuli in a native language* were investigated by means of a behavioural experiment. The results of the listening experiment showed that both hypotheses, proposed in Section 1.5 were rejected. While all of the stimuli, both Japanese and English, showed significant speech-to-song transforming scores, there was no significant main effect for *Native Language*. In line with the study by Jaisin et al. (2016), which showed a reduced STS illusion effect in tone language speakers compared to stress-accent language speakers, a reduced transforming speech-to-song effect was expected for the native Japanese speaking group than the native Hiberno-English speaking group. The hypothesized correlation between native speakers of a tonal language, and Japanese, a pitch-accent language, was posited according to the findings of a study by Sato et al. (2007), who show similar neural activations for native speakers of tone languages and Japanese in pitch processing. As the effect of *Native Language* proved to be insignificant in the current study, the proposed hypothesis is not supported.

The results regarding the effect of *Stimulus Language* on the transformation scores showed significance, with the English stimuli resulting in higher transformation scores than the Japanese stimuli. While no significant effect of *Stimulus Language* was posited, a significant interaction between the two independent variables *Native Language* and *Stimulus Language* was hypothesised, in line with empirical evidence showing a stronger transformation effect for stimuli in a non-native language (Margulis et al., 2015). The outcome of the analysis did not, however, yield in favour of this hypothesis, with no significant interaction between the aforementioned variables being reported.

Overall, while none of the predicted hypotheses were borne out, a main effect of *Stimulus Language* was found, with English stimuli showing significantly higher transformation scores than the Japanese stimuli across both groups. This result could be interpreted as reflecting a general propensity for the English stimuli, obtained from an established set of transforming stimuli, proven to result in high transformation scores (Tierney et al., 2013). Contrarily, the edited Japanese stimuli set has not been tested on a population prior to the current study. The acoustic correlates of the stimuli would potentially need to be reevaluated and tested on a larger cohort of participants in order to substantiate the set as an effectively transforming set of stimuli. An alternative explanation for the higher transforming English stimuli could potentially reflect the rhythmic features of the Japanese language, which will be further explored in Section 4.2.

4.1 Qualitative results

As mentioned in Section 2.3, participants took part in a questionnaire, asking about language background and musical experience, as well as an open question, enquiring into how the participants experienced the STS illusion, if at all. The qualitative results from this open question provided some interesting discussion points regarding the findings from the data analysis. Namely, two strong trends could be seen in the participant’s responses; the noting of the rhythmic features of the stimuli, as well as highlighting the higher likelihood of song-like transformations when the stimuli were presented in a non-native language. Tables (8) and (9) below provide some extracted examples, demonstrating these two trends.

TABLE (8) | *Questionnaire extracts from native English speaking group*

Extract from Native English Speakers	*Trend*
"I found it interesting how things began to sound more rhythmic after they were repeated."	Rhythmic features
"When it repeated, it sounded more musical but especially when it was Japanese."	Non-native transformation
"The repetition of the phrases in their rhythm appeared to be more song-like"	Rhythmic features
"The more I listened the more I picked up a song and repeated the beats"	Rhythmic features

TABLE (9) | *Questionnaire extracts from native Japanese speaking group*

Extract from Native Japanese Speakers [3]	*Trend*
"I experienced speech to song illusion, especially when I heard English sentences."	Non-native transformation
"felt rhythmic to me."	Rhythmic features
"English stimuli sounded more like a song than Japanese stimuli to me."	Non-native transformation
"Non-Japanese stimuli transformed upon repetition."	Non-native transformation
"When repeated 10 times, the English speech sounded more significantly like a song than Japanese."	Non-native transformation

The results regarding the rhythmic features did not elucidate whether either language showed a higher prevalence for this trend, however, the native English-speaking group did report this trend more frequently than the native Japanese-speaking group. This trend is of interest

[3] The questionnaire extracts from the Japanese speaking group were translated into English by a native Japese speaker.

when acknowledging the empirical evidence showing that rhythmic features of speech or a secondary mediator of this illusion, with the pitch being the highest predictor of a transformation (Tierney et al., 2018). With the English stimuli being taken from an established and well-attested transforming set (Tierney et al., 2018), it could be posited that the participant's answers regarding the rhythmic features could be largely in response to the Japanese stimuli set, which had previously not been tested in other studies investigating the STS illusion. These answers could potentially reflect the influence of a language's rhythmic features on how successfully their stimuli induce the perceptual transformation. This issue will be further explored in the following section.

The trend of non-native transformation was of interest with respect to the sub-research question, which posited a higher transformation for stimuli presented in a non-native language. The results of the data analysis showed, however, that this hypothesis was not borne out. The frequent mention of the non-native stimuli transforming more successfully than stimuli in a native language was particularly notable in the native Japanese-speaking group. Interestingly, from Figure (7) in Section 3.3, a high rating for non-native stimuli (Eng-JP and JP-Eng) can clearly be observed. This would indicate that due to an initial high song-like rating, the transformation scores were reduced, without showing low “song-like” scores for any of the non-native stimuli. This initial higher rating of non-native stimuli can also be seen in the study which motivated the sub-research question, by Margulis et al. (2015). In their results, there is a clear trend in participants' responses to the initial playing of the non-native languages, showing that the more dissimilar a language is to one’s native language, the higher the initial “song-like” score is.

The results of the current paper could thus reflect this finding by Margulis et al. (2015), or alternatively reflect the issue of stimuli matching across the English and Japanese stimuli. As mentioned, the Japanese stimuli set were not previously tested for the *speech-to-song transformation*, so it is possible that the inability to match acoustic features, such as pitch contour, across the two sets of stimuli, influenced the propensity for the Japanese stimuli to transform for both native English speakers, as well as native Japanese speakers. An overview of the tallied trend responses can be seen in Table (10).

TABLE (10) | *Tallied trends from quantitative results*

Group	*Mention of Rhythmic Features*	*Mention of Higher Transformation of Non-native Stimuli*
Native English Speakers	3	1
Native Japanese Speakers	1	4

4.2 Potential Influence of Language Specific Prosodic Features

As previously discussed, two of the main mediators of the *speech-to-song* transformation are pitch contour, being the most influential factor (Tierney et al., 2018), and regularly recurring temporal relations between accented and unaccented intervals acting as a secondary mediator (Falk et al., 2014). An interesting question in the investigation into the STS illusion is how specific acoustic features of a language can influence how the illusion is experienced for speakers of languages with differing prosodic features, such as rhythmic classification and pitch stability.

The rhythmic features of Japanese and English fall under two different categories of linguistic-rhythm classification; the former being *mora-timed,* the latter, *syllable-timed.* This distinction reflects how language systems split up recurring speech units, which can be intervals between stress units, rhythmic feet, syllables or morae (Grabe & Low, 2013). One of the most notable features of mora-timed languages is that they are characterized by each speech unit being equally timed (Tsujimura, 1999). This feature of Japanese proves of interest for the current study, as it appears to align with the rhythmic conditions necessary for the STS illusion to result in a transformation. As can be seen in Figure (6) in Section 3.3, the native English speakers showed a relatively high mean score for the pre-repetition Japanese stimuli, which in turn led to a reduced overall transformation effect. Considering the fact that the rhythmic features of a phrase act as secondary mediators of the STS illusion, with pitch contour acting as the primary mediator, it could be posited that, in the case that the rhythmic features of a phrase are more prominent in perception than pitch, that the overall transformation effect of the illusion is reduced. This could reflect the numerous open-question responses from native English speakers regarding rhythmic features, as well as their average initial high *song-like* scores, but overall low transformation scores.

Furthermore, the fundamental pitch contour differences in Japanese and English may have influenced the results of the current study. The issue with conducting a cross-linguistic study with the aim of comparing transformation scores which are mediated by particular acoustic features is that these features differ greatly from language to language. As mentioned, Japanese is both a pitch-accent language, using variation in F0 to denote lexical meaning, as well as consisting of near-equal duration intervals (Tsujimura, 1999) meaning that the rate of F0 change does not tend to follow the trajectory of a stable pitch contour. Considering both rhythmic and pitch features of Japanese discussed in this section, it could be posited that the rhythmic features resulted in initial high scoring of the Japanese stimuli by native English speakers, while the restriction in terms of pitch contour stability may have resulted in the overall low transformation scores.

Further research investigating the influence of language-specific acoustic features on the transformability of this illusion could potentially elucidate this seemingly complex relation between stimulus language and its perception across speakers of different languages. While the current empirical evidence suggests that a stable pitch contour is the most salient mediator of the illusion, this has only been shown to be reliable in speakers of a non-tonal language. Future cross-linguistic studies could investigate whether an alternative approach in terms of stimuli acoustic features, could provide more insight into how universal these proposed mediators of the STS illusion are.

4.3 Conclusion

To conclude, the present study adds to the gap in the literature regarding the efficacy of the STS illusion in native speakers of a pitch-accent language. While the results of the experiment did not align with the hypotheses, they did show a significant effect of Stimulus Language on the transformation. This unexpected outcome raised a number of interesting discussion points with regard to cross-linguistic work into the STS illusion. Namely, the question as to whether the current proposed mediators of the illusion, such as a stable pitch contour, can be generalised to speakers of differing language backgrounds. A further larger-scale, cross-linguistic study including stimuli from a constructed language, where acoustic features can be finely manipulated could potentially give further insight into this question. Furthermore, the findings presented in this study may reflect that the degree of tonal use in a language can mediate the effect of the speech-to-song illusion, resulting in a variance in the efficacy of this illusion between pitch-accent and tone language speakers. Future work comparing the illusion in the aforementioned groups would be necessary in order to understand the relationship between the role of pitch in a language and the efficacy of the STS illusion.

One challenge faced in the undertaking of this study was the matching of English and Japanese stimuli. While duration was matched as closely as possible, speaker's pitch range (reflected in stimuli from male and female speakers for the English stimuli, and only female for Japanese stimuli), and pitch contour were not controlled for. This variance in pitch properties may have influenced how comparable the two stimuli sets were, which may in turn have affected the experiment outcomes. While the issue of stimulus matching across languages was discussed in the previous section, future studies could look into how these aspects can be controlled and matched, without altering the naturalness of the stimuli.

In sum, this paper adds to the body of literature investigating this perceptual border between speech processing and song processing, adding to the research into how language background modulates the STS illusion. This contribution to the body of work on the STS illusion furthers the knowledge of the overlapping processes of language and music, offering further insight into how our mental representations across the two domains interact.

5. References

Bidelman, G. M., Hutka, S., & Moreno, S. (2013). Tone Language Speakers and Musicians Share Enhanced Perceptual and Cognitive Abilities for Musical Pitch: Evidence for Bidirectionality between the Domains of Language and Music. *PLoS ONE*, *8*(4), e60676. https://doi.org/10.1371/journal.pone.0060676

Boersma, P., & Weenick, D. (2023). Praat : doing phonetics by computer [Computer program]. *Glot International, Version 6.3.10*, retrieved April 2023 from https://www.praat.org

Castro, N., Mendoza, J., 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

Deutsch, D., Henthorn, T., & Lapidis, R. (2011). Illusory transformation from speech to song. *Journal of the Acoustical Society of America*, *129*(4), 2245–2252. https://doi.org/10.1121/1.3562174

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

Grabe, E., Kochanski, G., & Coleman, J. (2005). The intonation of native accent varieties in the British Isles: Potential for miscommunication. *English Pronunciation Models: A Changing Scene*, 311–337.

Grabe, E., & Low, E. L. (2013). Durational variability in speech and the Rhythm Class Hypothesis. In *De Gruyter eBooks*. https://doi.org/10.1515/9783110197105.515

Groenveld, G., Burgoyne, J., & Sadakata, M. (2020). I still hear a melody: investigating temporal dynamics of the Speech-to-Song Illusion. Psychological Research-psychologische Forschung, 84(5), 1451–1459. https://doi.org/10.1007/s00426-018-1135-z

Gussenhoven, C., & Gussenhoven, P. O. G. a. E. P. C. (2004). *The Phonology of Tone and Intonation*. Cambridge University Press.

Hyman, L. M. (2006). Word-prosodic typology. *Phonology*, *23*(02), 225–257. https://doi.org/10.1017/s0952675706000893

Hyman, L. M. (2009). How (not) to do phonological typology: the case of pitch-accent. *Language Sciences*, *31*(2–3), 213–238. https://doi.org/10.1016/j.langsci.2008.12.007

Jaisin, K., Suphanchaimat, R., Candia, M. a. F., & Warren, J. D. (2016). The Speech-to-Song Illusion Is Reduced in Speakers of Tonal (vs. Non-Tonal) Languages. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.00662

Margulis, E. H., Simchy-Gross, R., & Black, J. L. (2015). Pronunciation difficulty, temporal regularity, and the speech-to-song illusion. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.00048

Müllensiefen, D., Gingras, B., Musil, J., & Stewart, L. (2014). The Musicality of Non-Musicians: An Index for Assessing Musical Sophistication in the General Population. *PLoS One*, *9*(2), e89642. https://doi.org/10.1371/journal.pone.0089642

Patel, A. D. (2010). *Music, Language, and the Brain* (1st ed.). Oxford University Press.

Pfordresher, P. Q., & Brown, S. (2009). Enhanced production and perception of musical pitch in tone language speakers. *Attention, Perception, &Amp; Psychophysics*, *71*(6), 1385–1398. https://doi.org/10.3758/app.71.6.1385

Posit team (2023). *RStudio: Integrated Development Environment for R. Posit Software*, (Version 2023.3.1.446) [Software]. http://www.posit.co/.

Sadakata, M., Yamaguchi, Y., Ohsawa, C., Matsubara, M., Terasawa, H., Von Schnehen, A., Müllensiefen, D., & Sekiyama, K. (2022). The Japanese translation of the Gold-MSI: Adaptation and validation of the self-report questionnaire of musical sophistication. *Musicae Scientiae*, 102986492211100. https://doi.org/10.1177/10298649221110089

Sato, Y. S., Sogabe, Y., & Mazuka, R. (2007). Brain responses in the processing of lexical pitch-accent by Japanese speakers. *Neuroreport*, *18*(18), 2001–2004. https://doi.org/10.1097/wnr.0b013e3282f262de

Sato, Y. S., Utsugi, A., Yamane, N., Koizumi, M., & Mazuka, R. (2013). Dialectal differences in hemispheric specialization for Japanese lexical pitch accent. *Brain and Language*, *127*(3), 475–483. https://doi.org/10.1016/j.bandl.2013.09.008

Tierney, A., Dick, F., Deutsch, D., & Sereno, M. I. (2013). Speech versus Song: Multiple Pitch-Sensitive Areas Revealed by a Naturally Occurring Musical Illusion. *Cerebral Cortex*, *23*(2), 249–254. https://doi.org/10.1093/cercor/bhs003

Tierney, A., Patel, A. D., & Breen, M. (2018). Acoustic foundations of the speech-to-song illusion. *Journal of Experimental Psychology*, *147*(6), 888–904. https://doi.org/10.1037/xge0000455

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. https://doi.org/10.1037/xhp0000968

Tsujimura, N. (1999). The Handbook of Japanese Linguistics. In *Blackwell Publishing Ltd eBooks*. https://doi.org/10.1002/9781405166225

Vanden Bosch Der Nederlanden, C. M., Hannon, E. E., & Snyder, J. S. (2015). Everyday musical experience is sufficient to perceive the speech-to-song illusion. *Journal of Experimental Psychology: General*, *144*(2), e43–e49. https://doi.org/10.1037/xge0000056

Vet, D. J. (2023). *Experiment Designer* [Software], retrieved April 2023 from https://www.fon.hum.uva.nl/dirk/ed.php

Yip, M. (2002). *Tone*. Cambridge University Press.

Zatorre, R. J., & Gandour, J. T. (2007). Neural specializations for speech and pitch: moving beyond the dichotomies. *Philosophical Transactions of the Royal Society B*, *363*(1493), 1087–1104. https://doi.org/10.1098/rstb.2007.2161

6. Appendices

6.1 Information Brochure

6.1.1 English Information Brochure

Dear participant,

You will be taking part in the '*Speech or song?* Investigating the efficacy of the speech-to-song illusion in native speakers of a pitch-accent language' research project conducted by Martha Nobbe Smyth under supervision of Dr. M. Sadakata and Dr. B.M. van 't Veer at the University of Amsterdam, Department of Linguistics. Before the research project can begin, it is important that you read about the procedures we will be applying. Make sure to read this brochure carefully.

Purpose of the research project

When do you know whether somebody is speaking or singing? It may seem obvious; however, speech and song are remarkably similar in their acoustic features, both involving meaningful sequences and involving the human voice. A naturally occurring boundary between speech and song is highlighted via the so-called "speech-to-song illusion", a dramatic shift in the perception of speech to sound more song-like after multiple repetitions.

Over the course of this research, I aim to further investigate the phenomenon of the speech-to-song illusion, by observing how effective this phenomenon is across different languages. Previous research has shown that one's native language background does impact the extent to which the speech-to-song illusion can be elicited, providing interesting context for the current research project being undertaken. Data provided by participants via the listening experiment will enable the comparison of the efficacy of the perceptual illusion cross-linguistically, adding to the body of established work on this phenomenon. This will furthermore help us understand which features of speech lead to this illusion. I believe this research is of importance as it aims to offer insight into the overlap in processing between music and language, helping us to further understand how these two domains are cognitively linked.

Who can take part in this research?

We are inviting monolingual native English, and native Japanese speakers over the age of 16 years old to take part in this research. Before the experiment begins, you will be asked some questions about your hearing and eyesight. You may wear glasses, contact lenses or hearing aids. You will also be asked several questions about your musical experience and language background. You can take part in this research project if English is your mother tongue, and you were brought up in a bilingual household. We also need to make sure that you do not, to the best of your knowledge, have any language problems such as dyslexia or a specific language disorder.

Instructions and procedure

In order to participate in this experiment, you must have access to a device with working sound in order to listen to multiple media files, as well as a quiet environment. Headphone wearing is preferable. During the first part of the experiment, you will be asked a number of questions pertaining to your language background, musical experience and general information. The second part of the questionnaire will include a listening experiment, in which brief speech samples will be initially played once. You will then rate this on a 7-point scale, ranging from "more speech-like" to "more song-like". After registering your response, the same speech segment will be played multiple times, followed by the same 7-point scale rating, which you will again be asked to respond to. The total duration of the experiment will be 15-20 minutes.

Voluntary participation

You will be participating in this research project on a voluntary basis. This means you are free to stop taking part at any stage. This will not have any consequences and you will not be obliged to finish the procedures described above. You can always decide to withdraw your consent later on. If you decide to stop or withdraw your consent prior to publication of the research results, all the information gathered up until then will be permanently deleted. However, if information has been anonymized, it cannot be deleted because it is not possible to trace back the information to individual participants.

Discomfort, Risks & Insurance

The risks of participating in this research are no greater than in everyday situations at home. Previous experience in similar research has shown that no or hardly any discomfort is to be expected for participants. For all research at the University of Amsterdam, a standard liability insurance applies.

Confidential treatment of your details

The information gathered over the course of this research will be used for further analysis and publication in scientific journals only. Your personal details will not be used in these publications, and we guarantee that you will remain anonymous under all circumstances.

The data gathered during the research will be encrypted and stored separately from the personal details. These personal details and the encryption key are only accessible to members of the research staff. Anonymous data will be stored for a period of a minimum of 10 years. The personal data will only be stored as long as is necessary for the research and will be deleted as soon as possible.

Further information

For further information on the research project, please contact Martha Nobbe Smyth (email: martha.nobbe.smyth@student.uva.nl), Dr. M. Sadakata (email: m.sadakata@uva.nl) or Dr. B.M. van ‘t Veer (phone number: +31 20 525 38 72; email: b.m.vantveer@uva.nl; Spuistraat 134, 1012 VB Amsterdam, 6.38).

If you have any complaints regarding this research project, you can contact the secretary of the Ethics Committee of the Faculty of Humanities of the University of Amsterdam, commissie-ethiek-fgw@uva.nl, phone number: +31 20 – 525 3054; Binnengasthuisstraat 9, 1012 ZA Amsterdam.

6.1.2 Japanese Information Brochure

参加者の皆様
この度はアムステルダム大学の研究プロジェクト、「話し言葉か歌か？聴覚錯覚の実験」に興味を持っていただき、ありがとうございます。この研究はアムステルダム大学の貞方マキ子助教授、B.M. van 't Veer講師、及び東京女子大学の田中章浩教授の指導のもと、Martha Nobbe Smythによって行われます。はじめに、適用される手順について説明します。

研究プロジェクトの目的
聞こえてきた声が話しているのか歌っているのか、あなたにはすぐにわかりますか？話し声と歌は、一見全く違うように思えるかもしれませんが、実は両者の音響的特徴は非常に似ています。スピーチ・トゥ・ソング・イリュージョン（以下STS）と呼ばれる錯覚現象は、繰り返し聞くことで話し言葉が歌のように感じられるという、知覚の変化を指します。この錯覚は、話し言葉と歌が似ているからこそ起こるといえるでしょう。

本研究は、このSTSについて調査します。具体的には、この錯覚にもたらす母国語の影響を調べます。英語話者と日本語話者での錯覚の効果を比較し、言葉の特徴と歌の知覚がどのように関わっているのか理解を深めることが目的です。これにより、音楽と言語の処理がどの程度重複しているのか、さらにはこれら二つの領域が認知的にどのように結びついているかへの理解を深めることができると私たちは考えています。

参加資格
18歳から30歳までのバイリンガルではない英語または日本語の母語話者に参加をお願いしています。眼鏡、コンタクトレンズ、補聴器を使用しても構いません。また、あなたの音楽の経験と言語のバックグラウンドについてもいくつかの質問をします。失読症などの言語の障害を持つ方にはご参加いただけません。自分の参加資格に疑問がある場合にはご連絡ください。

手順
参加時には、音を聞くことができるコンピュータと、静かな環境が必要です。実験に際してはヘッドフォンの使用を強く推奨します。まず音が一度流れます。その後、あなたにその音がどのように聞こえたか、「話し言葉のよう」から「歌のよう」までの7段階で評価していただきます。回答を記録した後、同じ音が今度は繰り返し再生され、再度7段階の評価が求められます。この2回の評価で1セットになります。実験は最初に練習課題から始まり、続いて、本実験が始まります。最後にあなたの言語環境、音楽の経験、一般的な情報について質問をします。実験全体の所要時間は15〜20分です。

自由参加

あなたはこの研究プロジェクトに自由意志で参加しますので、いつでも終了することができます。終了することで、あなたが被る被害は一切ありません。また、いつでも実験参加の同意を

撤回できます。同意を撤回した場合、それまでに集められた情報はすべて削除されます。ただし、情報が匿名化されている場合は、個々の参加者に情報を遡ることができないため、削除することはできません。

リスク

この研究に参加するリスクは、家庭での日常的な状況と変わりありません。同様の研究の経験から、参加者にとって予想される不快感はほとんどまたは全くありません。アムステルダム大学が主催する全ての研究には責任保険が適用されます。

機密保持

この研究過程で収集された情報は、さらなる分析および科学雑誌での公表のためだけに使用されます。あなたの個人情報はこれらの出版物には使用されず、あらゆる状況下で匿名性は保証されます。研究中に収集されたデータは暗号化され、個人情報から切り離されて保管されます。これらの個人情報と暗号化キーは、研究スタッフのメンバーだけがアクセスできます。匿名のデータは最低でも10年間保管されます。個人情報は研究に必要な期間だけ保管され、可能な限り早く削除されます。

詳細情報

研究プロジェクトについての詳細情報は、Martha Nobbe Smyth（メール: martha.nobbe.smyth@student.uva.nl）、貞方マキ子（メール: m.sadakata@uva.nl、日本語でのメッセージ可能）またはDr. B.M. van 't Veer（メール: b.m.vantveer@uva.nl）にお問い合わせください。

苦情

万一研究プロジェクトに関して何か問題が発生した場合や、研究に関する情報が不十分だと感じた場合は、アムステルダム大学の倫理委員会（Ethics Review Board）に連絡することができます。メールアドレスは ethics-fmg@uva.nl です。
研究へのご協力に感謝申し上げます。

Martha Nobbe Smyth, M.Sadakata, B.M. van 't Veer & A. Tanaka

6.2 Informed Consent Form

6.2.1 English Informed Consent Form

‘I hereby declare that I have been clearly informed about the research project ‘*Speech or song?* A cross-linguistic comparison of the speech-to-song illusion’ at the University of Amsterdam, Department of Linguistics, conducted by Martha Nobbe Smyth under supervision of Dr. M. Sadakata and Dr. B.M. van ‘t Veer as described in the information brochure. My questions have been answered to my satisfaction.

I realise that participation in this research is on an entirely voluntary basis. I retain the right to revoke this consent without having to provide any reasons for my decision. I am aware that I am entitled to discontinue the research at any time, and that I can always withdraw my consent after the research has ended. If I decide to stop or withdraw my consent, all the information gathered up until then will be permanently deleted.

If my research results are used in scientific publications or made public in any other way, they will be fully anonymised. My personal information may not be viewed by third parties without my express permission.

If I need any further information on the research, now or in the future, I can contact Martha Nobbe Smyth (email: martha.nobbe.smyth@student.uva.nl), Dr. M. Sadakata (email: m.sadakata@uva.nl) or Dr. B.M. van ‘t Veer (phone number: +31 20 525 38 72; email: b.m.vantveer@uva.nl; Spuistraat 134, 1012 VB Amsterdam, 6.38).

If I have any complaints regarding this research, I can contact the secretary of the Ethics Committee of the Faculty of Humanities of the University of Amsterdam; email: commissie-ethiek-fgw@uva.nl; phone number: +31 20 – 525 3054; Binnengasthuisstraat 9, 1012 ZA Amsterdam.

I consent to:

- participate in this research yes / no

- my personal details to be stored for a period of 3 months yes / no

Signed in duplicate:

…………………………… ………………………. ……………………………

Name participant Date Signature

‘I have explained the research in further detail. I hereby declare my willingness to answer any further questions on the research to the best of my ability.’

…………………………… ………………………. ……………………………

Name researcher Date Signature

6.2.2 Japanese Informed Consent Form

同意書
私は、アムステルダム大学言語学部の研究プロジェクト「話し言葉か歌か？聴覚錯覚の実験」について、貞方マキ子、Dr. B.M. van 't Veer、田中章浩の指導のもと、Martha Nobbe Smythが実施することについて、十分な説明を受けたことを認めます。この実験に関して、疑問はありません。

私は、この研究への参加が完全に自由意志に基づくものであることを理解しています。私はこの同意を撤回する権利を保持し、その決定の理由を提供する義務はありません。いつでも研究を中止する権利があり、研究終了後もいつでも同意を撤回することができることを理解しています。中止するか、同意を撤回することにした場合、それまでに収集された情報はすべて永久に削除されます。

私の研究結果が科学的な出版物に使用されたり、他の方法で公開されたりする場合、それらは完全に匿名化されることを理解しています。私の個人情報を私の明示的な許可なしに第三者が閲覧することはありません。

もし今後研究についての詳細な情報が必要になった場合、Martha Nobbe Smyth（メール: martha.nobbe.smyth@student.uva.nl）、Dr. M. Sadakata（メール: m.sadakata@uva.nl、日本語でのメッセージ可能）またはDr. B.M. van ‘t Veer（メール: b.m.vantveer@uva.nl）に連絡を取ることができます。

研究について不備・不満があった場合には、アムステルダム大学人文学部倫理委員会の事務局に連絡することができます。メール: commissie-ethiek-fgw@uva.nl; 電話番号: +31 20 – 525 3054; Binnengasthuisstraat 9, 1012 ZA アムステルダム

クリックして実験に進むことで、私は以下に同意します：

- この研究に参加する
- 個人情報を3ヶ月間保管する

6.3 Stimuli

6.2.2 English Stimuli

English Stimil	***Duration (s)***	***Lowest Pitch***	***Highest Pitch***	***Time-Weighted Pitch Mean***	***SD***	***Pitch Range***
AT0001	1.44	98.5	229.4	162.7	26	130.9
AT0002	0.87	72.8	143.4	101.3	20.9	70.6
AT0003	0.78	76.4	124.4	97.5	11.7	48
AT0004	1.06	89.1	153.1	125.2	21.4	64
AT0005	0.97	191	365.5	264.5	52.2	174.5
AT0006	1.41	145.6	276.1	207.9	27	130.5
AT0007	0.99	108.2	171.5	142.6	17.3	63.3
AT0008	1.18	99.5	142.7	126.8	11.4	43.2
AT0009	1.4	72.2	133.7	99.9	15.3	61.5
AT0010	1.09	98.7	191	127.3	22.4	92.3
AT0011	0.84	96.9	156.5	117.7	15.1	59.6
AT0012	1.39	105.2	184.6	133.5	20	79.4
AT0013	1.29	108.7	350.8	153.1	48.1	242.1
AT0014	1.8	87.4	134.9	107.6	11	47.5
AT0015	1.14	172.1	283	231.1	25	110.9

6.2.2 Japanese Stimuli

Japanese Stimuli	***Duration (s)***	***Lowest Pitch***	***Highest Pitch***	***Time-Weighted Pitch Mean***	***SD***	***Pitch Range***
JP0001	0.911	190.5	418.3	255.2	68.6	227.8
JP0002	1.21	181	456.8	279.4	86	275.8
JP0003	1.46	177.4	344.5	251.8	38.4	167.1
JP0004	1.27	169	329.6	254.7	45	160.6
JP0005	1.54	148.8	346.7	251.1	66	197.9
JP0006	1.39	198	390.5	278.8	62.7	192.5
JP0007	1.94	192	611.3	276.3	79	419.3
JP0008	1.1	197.8	411.9	257.3	54	214.1
JP0009	1.47	186.5	426.7	260.6	64.3	240.2
JP0010	1.29	176.9	426	278	68	249.1
JP0011	1.36	211.9	544.4	317.9	72.4	332.5
JP0012	1.73	179.8	421.4	268	70.8	241.6
JP0013	1.46	179.7	382.5	266.2	50.4	202.8
JP0014	1.41	185.3	359.9	233.7	45.6	174.6
JP0015	1.37	170.2	482.2	328.5	112.2	312

6.4 Questionnaire

6.4.1 English Questionnaire

1. Age (in years)

2. Gender:

   ○ Male

   ○ Female

   ○ Other

3. Native Language(s)

   ○ English

   ○ Japanese

4. Which of the following dialects do you speak?

   ○ Hiberno (Irish) English

   ○ Other (specify)

5. Have you received any formal training in English?

   ○ Yes

○ No

i. (If "Yes”)Did you obtain an official certificate or diploma in one or more of the following?

○ Secondary School

○ JLPT

○ Other:

ii. How many years has it been since you finished your formal training in English?

iii. How would you rate your current proficiency in Japanese?

○ Beginner

○ Intermediate

○ Advanced

○ Near native

○ Native

iv. How often do you speak Japanese currently?

○ Daily

○ Weekly

○ Monthly

○ Less than once a month

v. - How often do you hear Japanese spoken currently?

○ Daily

○ Weekly

○ Monthly

○ Less than once a month

vi. How are you exposed to Japanese (more than one possible)?

1. ○ Japanese language music
2. ○ Film
3. ○ YouTube
4. ○ Social Media
5. ○ Japanese as language of instruction
6. ○ Other

6. Do you speak any second language, other than English?

7. I have never been complimented for my talents as a musical performer.

○ Completely Disagree

○ Strongly Disagree

○ Disagree

○ Neither Agree nor Disagree

○ Agree

○ Strongly Agree

○ Completely Agree

8. I would not consider myself a musician.

○ Completely Disagree

○ Strongly Disagree

○ Disagree

○ Neither Agree nor Disagree

○ Agree

○ Strongly Agree

○ Completely Agree

9. I engaged in regular, daily practice of a musical instrument (including voice) for (X) years

○ 0

○ 1

○ 2

○ 3

○ 4-5

○ 6-9

○ 10 or more

10. At the peak of my interest, I practiced (X) hours per day on my primary instrument.

○ 0

○ 0.5

○ 1

○ 1.5

○ 2

○ 3-4

○ 5 or more

11. I have had formal training in music theory for (X) years

○ 0

○ 0.5

○ 1

○ 2

○ 3

○ 4-6

○ 7 or more

12. I have had formal training of a musical instrument (including voice) for (X) years

○ 0

○ 0.5

○ 1

○ 2

○ 3-5

○ 6-9

○ 10 or more

13. I can play (X amount) musical instruments

○ 0

○ 1

○ 2

○ 3

○ 4

○ 5

○ 6 or more

14. What was your experience of the speech to song transformation if you experienced the speech-to-song illusion?

6.4.2 Japanese Questionnaire

1. 年齢:

2. 性別:

o 男性

o 女性

o その他

3. 母国語:

o 英語

o 日本語

4. あなたの日本語は標準語ですか？

o 標準語

o 方言

5. 英語の正式なトレーニングを受けたことがありますか（学校教育も含む）？

o はい

o いいえ

i. 以下の中から一つまたはそれ以上に公式の証明書または資格を取得しましたか？。

o 高校

o IELTS

o TOEFL

o その他:

ii. 英語の正式なトレーニングを終えてから何年が経ちましたか？

iii. 現在の英語の習熟度をどのように評価しますか？

o 初級

o 中級

o 上級

o ほぼネイティブ

o ネイティブ

iv. 現在、どの程度の頻度で英語を話しますか？

o 毎日

o 週に一回

o 月に一回

o 月に一度未満

v. 現在、どの程度の頻度で英語を聞きますか？

o 毎日

o 週に一回

o 月に一回

o 月に一度未満

vi. どのように英語に触れていますか（複数回答可能）？

o 英語の音楽

o 映画

o YouTube

o ソーシャルメディア

o 英語で授業を受けている

o 私の周りに英語を話す人がいる

o その他

6. 英語以外で日常会話がわかる程度以上に話せる言葉ありますか？それはどの言語ですか？（複数回答可）

7. 自分の音楽的才能や音楽演奏を褒められたことがない。

- o 全くあてはまらない
- o ほとんどあてはまらない
- o あまりあてはまらない
- o どちらでもない
- o わりとあてはまる
- o よくあてはまる
- o 実によくあてはまる

8. 私は音楽家ではない

- o 全くあてはまらない
- o ほとんどあてはまらない
- o あまりあてはまらない
- o どちらでもない
- o わりとあてはまる
- o よくあてはまる
- o 実によくあてはまる

9. 私はここ____年間、日常的・定期的に楽器（歌を含む）の練習をしている

○ 0

○ 1

○ 2

○ 3

○ 4-5

○ 6-9

○ 10かそれ以上

10. 興味が最も高かった時期には、メインの楽器（歌を含む）を一日に____時間練習した

○ 0

○ 0.5

○ 1

○ 1.5

○ 2

○ 3-4

○ 5かそれ以上

11. 音楽理論の訓練を____年間受けた

○ 0

○ 0.5

○ 1

○ 2

○ 3

○ 4-6

○ 7 かそれ以上

12. これまでに楽器（歌も含む）の正式な訓練を____年間受けた

○ 0

○ 0.5

○ 1

○ 2

○ 3-5

○ 6-9

○ 10 かそれ以上

13. 私は____種類の楽器を演奏できる

○ 0

○ 1

○ 2

○ 3

○ 4

○ 5

○ 6かそれ以上

14. この実験中に「繰り返し聞くことで話し言葉が歌のように感じられる知覚の変化」を体験しましたか？あなたの、\n実験中の音の知覚について、どんなことでも良いので気づいたことを教えてください