 doi:10.20944/preprints201808.0522.v1

Article

Pace, Emotion, and Language Tonality on Speech-to-song Illusion

Carole Leung [1*], De-Hui Ruth Zhou [2]

[1] Department of Counselling and Psychology, Hong Kong Shue Yan University; caroleleung69@gmail.com
[2] Department of Counselling and Psychology, Hong Kong Shue Yan University; dhzhou@hksyu.edu
* Correspondence: caroleleung69@gmail.com

Abstract: The speech-to-song illusion is a type of auditory illusion that the repetition of a part of a sentence would change people's perception tendency from speech-like to song-like. The study aims to examine how pace, emotion, and language tonality affect people's experience of the speech-to-song illusion. It uses a between-subject (Pace: fast, normal, vs. slow) and within-subject (Emotion: positive, negative, vs. neutral; language tonality: tonal language vs. non-tonal language) design. Sixty Hong Kong college students were randomly assigned to one of the three conditions characterized by pace. They listened to 12 audio stimuli, each with repetitions of a short excerpt, and rated their subjective perception of the presented phrase, whether it sounded like a speech or a song, on a five-point Likert-scale. Paired-sample *t*-tests and repeated measures ANOVAs were used to analyze the data. The findings reveal that a faster speech pace could strengthen the tendency of the speech-to-song illusion. Neither emotion nor language tonality show a statistically significant influence on the speech-to-song illusion. This study suggests that the perception of sound should be in a continuum and facilitates the understanding of song production in which speech can turn into music by having repetitive phrases and to be played in a relatively fast pace.

Keywords: speech-to-song illusion, auditory illusion, perception, pace, emotion, language tonality

1. Introduction

Biologically, speaking and singing are two similar processes. When people speak or sing, larynx, which is an organ located above the trachea, is involved in the production of those voices [1]. Once people listen to speech or song, there is the involvement of cochlea and temporal lobe of our brain for sound processing [2].

These two forms of communication are also similar in a way that people can express their thoughts and feelings through words. Stanley (1945) has regarded song as a "magnification of speech, which is wedded to the language of music" [3] (p. 268). For instance, rap music is a type of music that connects elements of singing and talking and it is popular among African Americans for the expression of fear, frustration, anger, etc. [4,5]. Likewise, performers in a musical sing out a story to the audiences.

Although speech and song are similar, people can easily categorize them since the melodic features and a structured rhythm defined different types of song from speech [6]. In recent years, however, the boundary between speech and song is found to be blurred. A song produced by the Chainsmokers named "#Selfie", which contains mainly monologue with upbeat electronic instrumental music, ranked top in Billboard's dance/electronic digital songs charts in 2014 [7]. In 2011, Deutsch et al. discovered the speech-to-song (STS) illusion, in which people's perception of speech would change to perception of song when a small part of the presented speech has repeated for 10 times [6].

By definition, perception is the process that people recognize, organize, and interpret sensory information received from the environment [8]. The blurring boundary between speech and song recently recognized leads to a scientific question about how people perceive a sound as speech or

 doi:10.20944/preprints201808.0522.v1

song. The present research aims at investigating how pace, emotion, and the language tonality of the speech affect people's perception and lead them to experience the STS illusion.

1.1 Perception of Speech and Song

Pace, pitch, and rhythm are the three key components of speech and music. Pace (in music, it is called tempo; in speech, it is named as speech rate) has an important role in the expression of emotions. If the speech rate is fast or if the tempo of a piece of music is fast, it could facilitate the expression of anger, fear, or happiness; if the pace of the speech or music is slow, it could convey feelings of sadness (Patel, 2008, as cited in [9]). Pitch refers to the listener's perceptual experience of the relative height of sounds [10]. It creates melody in music and prosody in languages [11]. Rhythm is the structured patterning of sounds in its pitch and accents [12].

Still, it is usually easy for people to distinguish between a speech and a song as they have slight differences in acoustical characteristics. In speech, there are rapid changes in loudness and pitch. Especially in non-tonal languages that is intonational, the contours of the pitch are broadly defined [6,13]. Moreover, compared with singing, people's pitch during the time when they are speaking is much lower. This helps to convey their extremely deep and heartfelt feelings [3].

In contrast, songs consist of discrete musical notes. Its pitch relationship as well as its rhythmic structure are more stable and well-defined [6]. Also, the singing voice can be sustained and moved from one tone to another without stopping or jerking by using the technique of vibrato, where this would seldom occur in speaking [3].

1.1.1. Pace and Music

In defining the pace of a piece of music, previous research mostly classified slow-tempo music as having a pace below 120 BPM (beats per minute) and fast tempo music as having a pace above 120 BPM [14-17].

The cognitive effect of fast-tempo music is different from the effect of slow-tempo music. For example, fast-tempo background music, in comparison to playing slow-tempo background music in an advertisement, was found to distract a person's cognitive resources from processing the content of the advertisement, thereby lowering the content recall rate [14].

Likewise, music with different tempos can affect human behaviors.

Milliman's study (1982) discovered that shoppers tend to walk faster under fast tempo background music; conversely, they walk slower under slow tempo background music [15]. When fast tempo music was played at a cafeteria, people consumed food faster, but an opposite effect of slow tempo music was not found [16]. Other behaviors, such as reading (Kallinen, 2002, as cited in [17]) or drawing (Nittono, Tsuda, Akai, & Nakajima, 2000, as cited in [17]), would have an increase in pace when fast tempo music was played.

1.1.2. Pace and Speech Perception

How fast or slow a speech was spoken could influence people's perception of the speech. A speech with normal speaking rate is often compressed or expanded to manipulate different paces of the stimuli for experimental use. Viewing several research studies, the time compression factor for a faster speech was usually at the range from 0.38 to 0.60, whereas for a slower speech, the time expansion factor was between 1.75 and 1.90 [18-20].

According to Dilley and Pitt (2010), listeners reported less function words (e.g. are, or, a) when the context speech rate has slowed down, but they heard a never-spoken function words when the context speech rate has speeded up [18]. A similar study conducted by Morrill et al. (2015) suggests a lower recall rate of sentences with a slower speech rate, compared with the original speech rate [20]. In another study, it was demonstrated that the effect of speech rate became stronger when the experimental session was extended to an hour [21]. Also, Bosker (2017) found that the speech rate of the context sentence could contribute to people's perception biased towards a particular target vowel

 doi:10.20944/preprints201808.0522.v1

[22]. Nevertheless, a study carried out by Dupoux and Green (1997) showed that people could perceptually adjust to the compressed utterance by having more exposure to it and this resulted in an improvement in recalling the sentences [19].

In addition to sentences, the pace of a word presented could also influence people's perception. Verbal transformation effect is a type of auditory illusion in which people heard words with similar sounds as the target word that are not presented when the target had been played again and again [23]. For example, people reported hearing "stress", "tress", or "Esther" when there was a repeated presentation of the word "rest" [23]. The effect would be reduced when the pace of the presented target word slows down [24].

1.1.3. Emotion

The emotion we feel influences our perception. In one of the experimental studies, Avramova, Stapel, and Lerouge (2010) used the Ebbinghaus illusion task to examine how emotions influence people's size judgment [25]. The Ebbinghaus illusion is a type of visual illusion where the central circle surrounded by small circles looks larger than another central circle surrounded by large circles, even though the two central circles are of the same size. Compared with participants in negative moods, those in positive mood made more inaccurate judgement on sizes. It was explained that people in the positive mood would attend to both the target and the context information, while people in the negative mood focus primarily on the target. Another experiment also provided evidence that positive emotions facilitate global processing of visual stimuli [26].

Besides visual perception, our auditory perception could be affected by emotion. A perceptual map has been yielded by Bergman, Västfjäll, Tajadura-Jimenez, and Asutay (2016) illustrating the role of emotion in perceiving and categorizing everyday sounds [27]. Existing literature show how emotions may affect the perception of loudness. People were found to rate tones louder under negative emotion condition than under the neutral condition [28,29].

1.1.4. Speech-to-song Illusion

Perceptual illusion is about people's perception of sensory information that may not present physically in the stimuli [8]. In particular, auditory illusion is related to the hearing and perception of sound. People may have reorganized the sound segments in mind which results in an alteration in the experience of the sound, so they may hear a sound that has not been played or they may not be able to report a sound that has actually been presented.

In hearing words, as mentioned above, the verbal transformation effect is one type of auditory illusion that the repetition of words could create words with similar sound as the presented target word [23]. There is also a mismatch between the stimulus and people's percept in hearing musical notes. When simultaneously and repeatedly play two tones that are spaced apart by an octave in alteration in both ears, listeners reported hearing one single tone that switched from ear to ear. This is known as the octave illusion [30]. Recently, a type of auditory illusion is found to be related to the perception of speech and song. It is called the speech-to-song (STS) illusion.

When people hear a short phrase repeatedly, their perception transformed from speech to song [6]. It is important to note that only participants' perception has been altered, the stimuli has not changed at all. In Deutsch et al. (2011)'s experiment, a sentence that lasted for 10.146 s was presented to the participants [6]. A short excerpt of 2.292 s was then played for 10 times. Following that, participants heard the original full sentence again. Right after each presentation, there was a 2300 ms pause where participants had to rate their perception of the phrase on a five-point Likert-scale with 1 representing "exactly like speech" and 5 representing "exactly like song". Comparing the mean ratings of the short excerpt on the first and final presentations, the rating score increased from around 1.30 to around 3.70.

This illusion occurred in everyone, no particular group of people are exempted. In spite of the fact that musically trained people have higher sensitivity to the pitch structure in music and language [31], both musically trained participants and casual music listener experienced the illusion [13].

 doi:10.20944/preprints201808.0522.v1

Neurological evidence revealed that the phrases presented had indeed activated several song perception related brain areas for pitch patterns processing and auditory motor processing, such as the anterior superior temporal gyrus, the right lateral precentral gyrus, and the left inferior frontal gyrus [32].

In fact, the tonality and familiarity of the presented speech can determine the occurrence of the illusion.

Language tonality. There are two main categories of the tonality in languages, namely, tonal language and non-tonal language. Tonal language is defined by its discrete pitch value of the syllable with lexical semantic meaning encoded [33,34]. Cantonese is a type of tonal language. It consists of six tones that defines the semantic meaning of the words. The six tones are high level (T1), high rising (T2), mid level (T3), low falling (T4), low rising (T5), and low level (T6) [35]. Take the monosyllable /tin/ as an illustration, it could mean sky when it is in a high level tone (/tin1/, 天) or it could mean farmland when it is in a low falling tone (/tin4/, 田). Unlike tonal language, non-tonal language is defined by its variation in pitch that would not affect the lexical meaning of the words [33,34]. In English, there are no specific tone for a particular word. The variations in pitch are primarily used for giving emphasis, expressing feelings, and indicating whether it is a statement or a question [9]. Changing the pitch of the words would not alter the meaning expressed.

The STS illusion was observed in both tonal language stimuli and non-tonal language stimuli. Among different tonal languages, Thai and Mandarin had been investigated [34]. The mean change of rating scores in participants' perception from speech to song was the same in these two tonal languages, which was 0.40. Contrary to tonal languages, there were slight differences in the mean rating change in several non-tonal languages examined. In Margulis, Simchy-gross, and Black (2015)'s experiment, Irish had the largest change in rating (mean rating change = 0.92) and Portuguese had the least change (mean rating change = 0.25) [36].

Even though the illusion occurs in all languages investigated, a difference was found between native tonal language speakers and native non-tonal language speakers on their perception of the spoken excerpt. As exemplified by Jaisin et al. (2016), the illusory change from speech to song was less obvious among people with mother tongue of a tonal language [34]. It was suggested that this group of people may have a tendency of not rating the stimuli as song-like due to the fact that they turn to code pitch patterns as linguistic.

Language familiarity. According to Margulis et al. (2015), the illusion occurred significantly stronger when the phrase presented was spoken in an unfamiliar language with difficult pronunciation [36]. Jaisin et al. (2016) also mentioned that not being able to comprehend the linguistic information of the utterance might facilitate the person to interpret its prosodic features to be musical [34]. Margulis et al. (2015) explained that the semantic satiation effect has given rise to the STS illusion. For this reason, the person would focus more on listening the sound of the phrase instead of its semantic meaning after hearing several repetitions [36].

1.2 Purposes

Even though a number of research has identified that the tempo in music could affect people's behavior (e.g. speed of walking) and memory (i.e. content recall rate) [14-17] as well as the speech rate could affect people's speech perception [18-22,24], particularly in the recognition of words, no existing study has ever tested the effect of pace on the auditory illusory transformation. This study is a pioneering study to investigate the effect of pace in affecting the perception of the STS illusion.

Regarding emotion, previous research discovered that the emotion people are experiencing could affect how they perceive visual information [25,26]. Nevertheless, the effect of emotion on people's auditory perception of speech and song has not been examined before. Hence, the second purpose of the experiment was to test the effect of emotion on the experience of STS illusion.

In addition, it is known from existing research that all languages could produce the STS illusion [6,13,34,36]. However, no previous study has ever focused on the tonal and non-tonal language effect with participants familiar with all the stimuli languages. The participants in Jaisin et al. (2016)'s and

 doi:10.20944/preprints201808.0522.v1

Margulis et al. (2015)'s experiment did not understand some of the stimuli languages at all [34,36]. Also, Cantonese, a type of tonal language with six contrasting tones, has not been studied before. The current experiment compared the effect of the STS illusion among Hong Kong college students in the two types of languages that they know well, which are Cantonese (i.e. tonal language) and English (i.e. non-tonal language).

The research question of this study is: How do pace, emotion, and language tonality affect STS illusion?

The hypotheses are:

H1: Stimuli played in a fast pace would lead to a stronger STS illusion.

H2: Positive emotion induced in the stimulus would lead to a stronger STS illusion.

H3: Non-tonal language would lead to a stronger STS illusion.

2. Materials and Methods

2.1 Design

This study used a between-subject and within-subject experimental design. As illustrated in Figure 1, the independent variables include pace, which was a between-subject variable, whereas the other two independent variables, emotion and language tonality, were the within-subject variables. Since Cantonese is the native language used in Hong Kong, it is the tonal language selected for investigation in this experiment. The non-tonal language selected is English. It is the second language commonly-used in Hong Kong. The dependent variable in this experiment was people's perception of speech and song, which was measured by their pre-repetition score and post-repetition score.

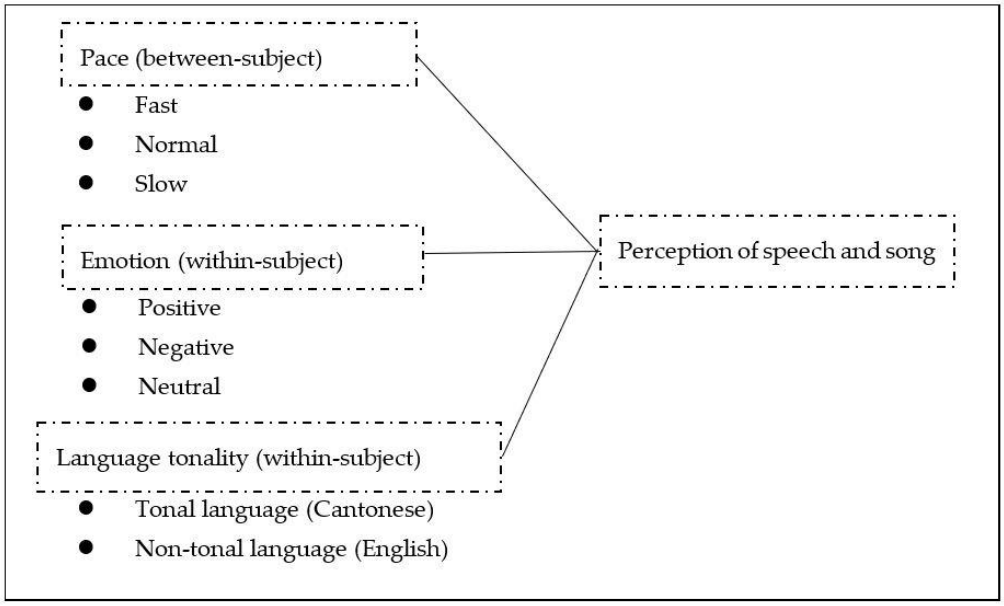


Figure 1. The design of the experiment.

2.2 Participants

Sixty local university students (16 males, 44 females; mean age = 20.18, *SD* = 1.40) were recruited in this study. None of them have hearing impairments, congenital amusia, or perfect pitch ability They were native Cantonese speakers and has been learning English since they were in kindergarten (mean age of start learning English = 3.61, *SD* = 1.08). It is also common among local universities students to speak more than two languages, both tonal languages (e.g. Cantonese, Mandarin) and non-tonal language (e.g. English, French). The median of the number of languages they can speak was 3. The average years of music training was 5.24 years (*SD* = 4.78; range: 0 – 16 years).

 doi:10.20944/preprints201808.0522.v1

The data sets of four other participants were excluded from the data analysis either because they did not meet the English language requirement of obtaining level three in English in the Hong Kong Diploma of Secondary Education Examination or other English language equivalent qualification (n = 3) or because they reported having cognitive amusia (n = 1).

Participants were randomly assigned to one of the three conditions characterized by the pace of the stimulus, which were the fast pace condition, the normal pace condition, and the slow pace condition. There were 20 participants in each condition.

2.3 Materials

2.3.1. Stimuli

A pilot test was carried out to prepare sentences with different emotionality. Twenty-three participants at the age of 18 to 25 rated 60 sentences extracted from books (30 Chinese sentences and 30 English sentences) on a five-point Likert scale with 1 representing "I feel strongly negative" and 5 representing "I feel strongly positive". The selection of stimuli was based on the scoring of the sentences (i.e. positive sentences: not less than a mean score of 4; negative sentences: not greater than a score of 2; neutral sentences: a score of approximately 3).

A total of 12 sentences (6 Chinese sentences and 6 English sentences) were then selected as the stimuli for the experiment. Within each language, two of the stimulus sentences convey positive emotions, another two of them convey negative emotions, and the other two were neutral sentences that consist of facts. Table 1 shows the mean rating score of emotion of the 12 sentences. Both positive Chinese and English sentences received a mean score of 4.13 (*SD* of Chinese sentences = .69; *SD* of English sentences = .71). For negative emotion, the Chinese sentences had a mean score of 1.63 (*SD* = .41) and the English ones had a mean score of 1.93 (*SD* = .43). The mean score of the neutral Chinese sentences and the neutral English sentences were 3.48 (*SD* = .49) and 3.11 (*SD* = .23) respectively.

Table 1. The mean rating score of emotion of the 12 sentences.

Emotion	Language	Mean Rating Score	*SD*
Positive	Chinese	4.13	.69
	English	4.13	.71
Negative	Chinese	1.63	.41
	English	1.93	.43
Neutral	Chinese	3.48	.49
	English	3.11	.23

All the sentences were recorded by a female speaker speaking in a normal rate, in which six of them were spoken in Cantonese and the rest were spoken in English. In fact, there are two subtypes of Cantonese, the colloquial form and the written form. Colloquial Cantonese is primarily used informally in daily causal conversation; whereas written Cantonese is formally used in newspaper, textbooks, documents, etc. For instance, the phrase "你想做什麼?" ("What are you going to do?") is in written form and its colloquial form is "你想做咩?" ("Whatcha gonna do?"). To simulate Cantonese songs in the present experiment, the stimulus sentences were spoken in the written language form of Cantonese since most Cantonese songs are sung in written language rather than in colloquial language. Table 2 below displays some of the stimulus sentences that were used for the experiment.

Table 2. Sample of the stimuli that were used.

Types	Full sentence	Length	Excerpt		
			Content	Length	No. of syllables
Positive emotions	I always found peace and comfort gazing at the mountains or watching the sun set on the beach in my new environment.	8.090 s	I always found peace and comfort	2.066 s	8
	我總是躺在草蓆上，一分一秒的等候著黃昏的來臨，那時候，只有黃昏涼爽的風來了，使我能在門外坐一會，就是我所期盼著的最大的幸福了。	15.895 s	我總是躺在草蓆上	2.344 s	8
Negative emotions	Deep down, I knew it wouldn't go away. I held her again, not knowing what else to do, tears filling my eyes, trying and failing to be the rock I think she needed.	11.448 s	Deep down, I knew it wouldn't go away	2.203 s	10
	我們的心碎了。眼看著一個個人死去，耳聽著一聲聲呼救，我們直淌眼淚，毫無辦法。甚麼也沒有啊!	12.912 s	我們的心碎了	1.605 s	6
Neutral	Perception is the set of processes by which we recognize, organize, and make sense of the sensations we receive from environment stimuli.	9.916 s	Perception is the set of processes	2.418 s	10
	中國最崇高的理想，就是一個人不必逃避人類社會和人生，而本性仍能保持原有的快樂。	9.626 s	中國最崇高的理想	1.990 s	8

As similar in previous research [6,13,34,36], the mean length of the full stimulus sentences was 11.51 s (*SD* = 2.06) and the mean duration of the short excerpt was 2.06 s (*SD* = .42) with number of syllables between 6 and 10. All short excerpts were extracted by using Audacity 2.1.3 [37].

The pace of the stimuli was manipulated by PSOLA in Praat software [38]. The time-compression factor was 0.65 for the fast pace condition. The time-expansion factor was 1.65 for the slow pace condition. All recordings were saved as WAV signed 16-bit PCM file at a sampling frequency of 44.1 kHz.

2.3.2. Questionnaire and measures

A questionnaire was designed to understand participants' demographic information and to make sure all participants are bilingual with Cantonese as their mother language and do not have hearing difficulties, cognitive amusia, or perfect pitch ability.

The present study followed prior experiments in measuring people's perception of how the stimulus phrases sound like, whether it is perceived as speech-like or song-like [6,13,34,36]. A five-point Likert scale was used for measurement, where 1 represents "exactly like speech" and 5 represents "exactly like song". Participants' responses were recorded by clicking the number of the Likert-scale showed on the computer screen.

2.4 Procedures

Participants signed the informed consent before taking part in the experiment. They were randomly assigned to the fast, the normal or the slow pace condition. Procedures for all conditions are the same, only the pace of the stimulus sentences are different. Figure 2 presents the procedures of this experiment.

 doi:10.20944/preprints201808.0522.v1

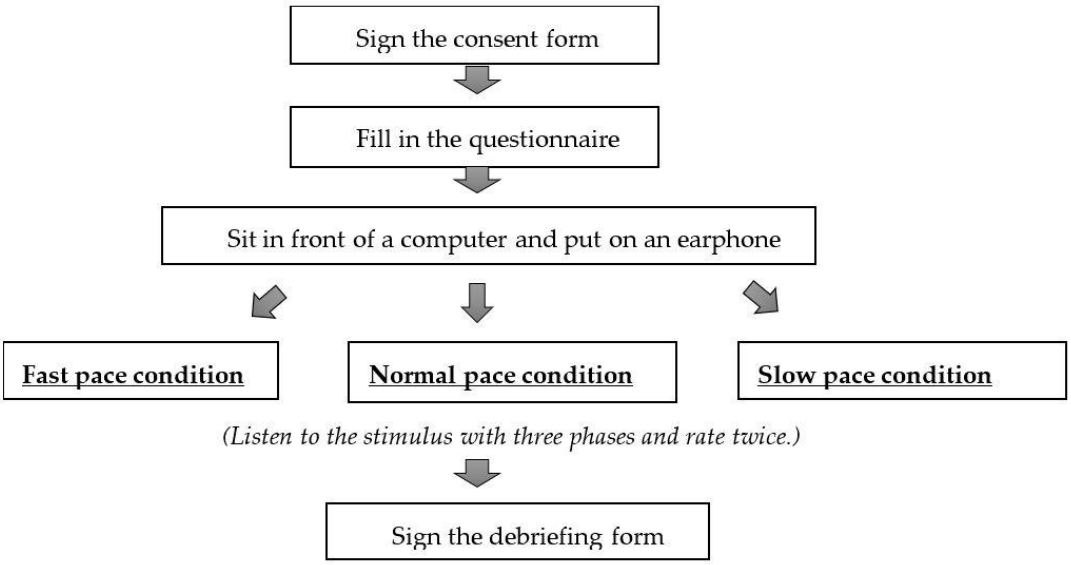


Figure 2. A flow chart showing the procedure of the experiment.

Participants were seated in front of a computer (screen size 16.8 x 29.5 cm) in a quiet place at campus and were required to put on an earphone. The distance between them and the monitor screen was approximately 40 cm. Google form was used as an online experimental platform. All of their responses were made via the use of the keyboard and the mouse.

First of all, the instructions of the experiment were displayed on the screen. Participants had to click the continued button by themselves at the end of the page when they are ready. Following the experimental procedures of previous studies [6,13,34,36], there are three consecutive phases for the stimulus. In the first phase, participants were asked to listen carefully to the full sentence once (i.e. Deep down, I knew it wouldn't go away. I held her again, not knowing what else to do, tears filling my eyes, trying and failing to be the rock I think she needing). In the next phase, the selected short excerpt (i.e. Deep down, I knew it wouldn't go away) was then played repeatedly for 10 times with a pause of a duration of 2300 ms in between each repetition. This was followed by playing the full sentence once again in the last phase. Figure 3 shows the arrangement of the stimulus.

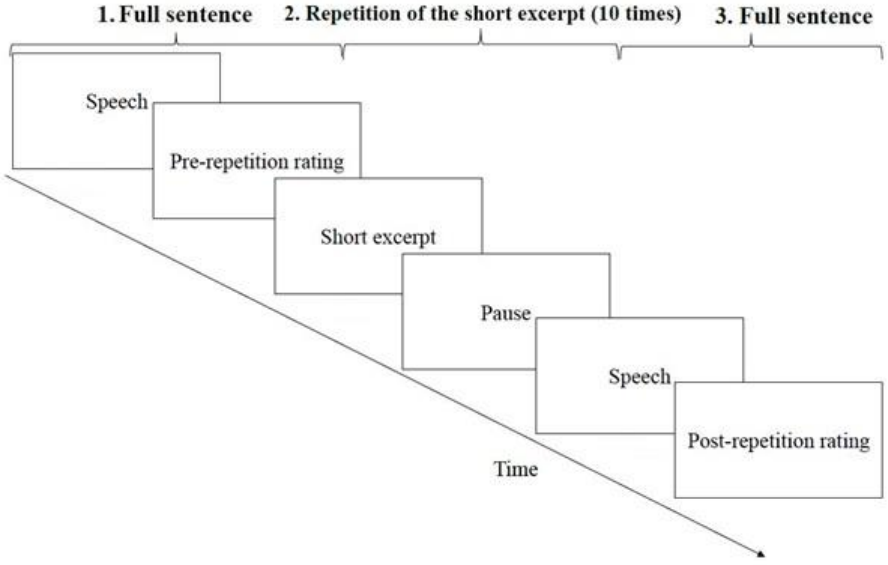


Figure 3. The arrangement of the three phases of the stimulus. The words of the stimulus were not shown to the participants.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 August 2018 doi:10.20944/preprints201808.0522.v1

For each stimulus, participants had to rate twice, once after the first presentation of the full sentence in the first phase and once after the last phase. Participants were asked to rate their subjective perception of how they think the presented phrase has sounded like immediately by using a mouse to click on a number in the five-point Likert scale displayed on the computer screen (see Figure 4). They were reminded not to rate the phrase based on what it originally is, but rather, what they feel the phrase is.

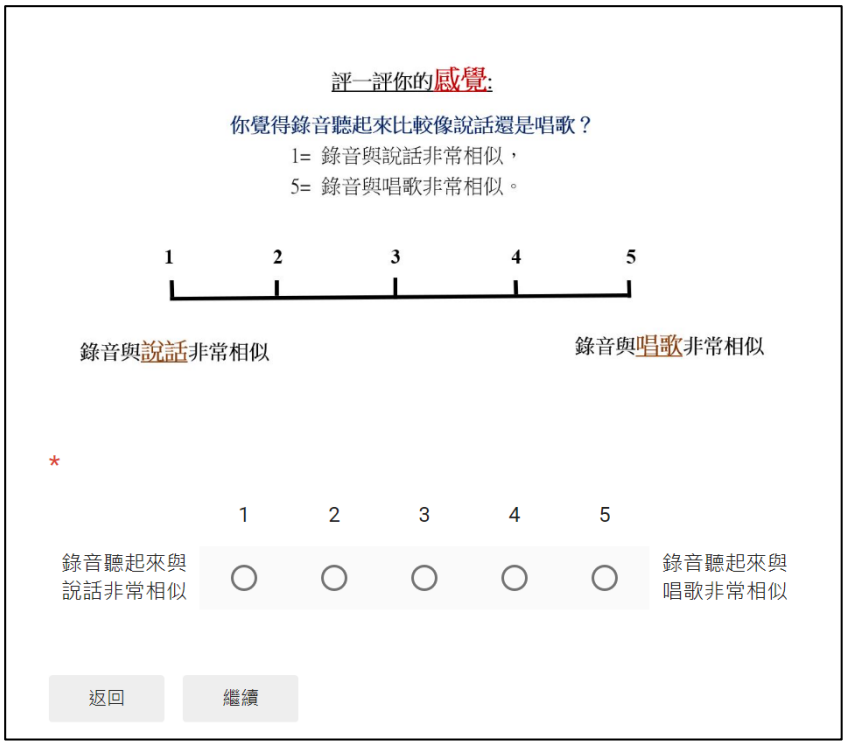


Figure 4. The computer screen showing a five-point Likert scale.

The experiment lasted for 25 minutes. During the experimental process, participants listened to 12 trials with both languages included. The trials were played in randomized order. 22 sets of 12 numbers per set were created by research randomizer [39]. After the completion of the experiment, all participants were given a debriefing form.

2.5 Research ethics

Before the start of the experiment, all participants are required to sign a consent form. In the experimental process, they may experience fatigue after listening to several trials, each with 10 repetitions of the phrases. Yet, this fatigue is tolerable. Participants can ask for a break and click on the continued button to listen to the next trial until they are ready. If participants cannot stand the fatigue, they could choose to discontinue their participation at any time during the experiment. After they completed the experiment, they were debriefed about the purposes of the study. The principles and procedures used in the study concerning human research ethics were approved by the Human Research Ethics Committee of Hong Kong Shue Yan University in August 2017.

3. Results

3.1 The Speech-to-song Illusion

Paired sample *t*-tests were computed to compare the significant difference between the pre-repetition rating and the post-repetition rating. To examine if the groups were significantly different from each other, two repeated measure ANOVAs were performed to compare the mean rating

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 August 2018 doi:10.20944/preprints201808.0522.v1

change within or between groups. The ANOVAs were 2 (ratings) x 3 (emotion) ANOVA, and 2 (ratings) x 2 (language tonality) ANOVA.

On the whole, the post-repetition rating (M = 1.95, SD = .70) of the stimuli was significantly higher than the pre-repetition rating (M = 1.71, SD = .54), $t(59) = 3.23$, $p < .01$, $d = 0.394$. This indicated a small change in perception from perceiving the stimuli as speech to perceiving the stimuli as more song-like after the repetition of the excerpts.

3.1.1. The Effect of Pace

Analyzing each condition separately, it was found that the STS illusion only occurred when stimuli were played in a fast pace, which had a rise of score from a mean pre-repetition rating of 1.800 (SD = .48) to a mean post-repetition rating of 2.217 (SD = .70), $t(19) = 3.055$, $p < .01$, $d = 0.696$ (see Table 5). In spite of the fact that the post-repetition rating score was higher than the pre-repetition score in normal pace (Pre-repetition score: M = 1.76, SD = .66; Post-repetition score: M = 1.82, SD = .60) and slow pace (Pre-repetition score: M = 1.56, SD = .47; Post-repetition score: M = 1.83, SD = .75), paired sample t-tests showed that the two rating scores did not differ significantly from each other. Hence, the first hypothesis that fast pace lead to a stronger STS illusion is confirmed. When self-selected stimulus was used, only those played in a fast pace could give rise to the illusory transformation.

3.1.2. The Effect of Emotion

As illustrated in Table 5, all three types of emotion could significantly result in the STS illusion. Stimuli in positive emotion had an elevation in rating from 1.99 (SD = .67) to 2.28 (SD = .80), $t(59) = 3.258$, $p < .01$, $d = 0.399$. Stimuli in negative emotion had a rise in rating from 1.74 (SD = .65) to 1.95 (SD = .82), $t(59) = 2.475$, $p < .05$, $d = 0.296$. The ratings increased from 1.45 (SD = .50) to 1.62 (SD = .71) in neutral stimuli as well, $t(59) = 2.199$, $p < .05$, $d = 0.286$.

Yet, no one particular type of emotion could cause a stronger illusion. The analysis from ANOVA showed no significant interation effect between the two ratings and the three types of emotion, $F(2, 58) = 1.431$, $p > .05$, $\eta_p^2 = .047$. The mean rating change among the three groups characherized by emotion (positive, negative, neutral) did not differ, the second hypothesis that positive emotion induced in the stimuli would lead to a stronger STS illusion is therefore rejected.

3.1.3. The Effect of Language Tonality

As seen in Table 5, both Cantonese and English could significantly result in the STS illusion. There was an increase from 1.56 (SD = .56) to 1.78 (SD = .71), $t(59) = 2.749$, $p < .01$, $d = 0.330$, when stimuli were spoken in Cantonese. An increase in ratings was also observed in stimuli spoken in English. The score rose from 1.85 (SD = .66) to 2.13 (SD = .81), $t(59) = 3.097$, $p < .01$, $d = 0.384$.

The ratings x language tonality interaction was not significant, $F(1, 59) = 1.018$, $p > .05$, $\eta_p^2 = .017$. Between Cantonese (tonal lanugage) and English (Non-tonal language), the change of the mean rating were no different from each other. So, the third hypothesis that non-tonal language would lead to a stronger STS illusion is not supported.

Table 3. The Mean Rating Score and the Paired Sample T-Tests Comparison of the Score.

IV	Levels	N	Rating	Mean	SD	Mean rating change	t	d
Pace	Fast	20	Pre-repetition	1.80	.48	.42	3.055**	0.696
			Post-repetition	2.22	.70			
	Normal	20	Pre-repetition	1.76	.66	.06	.468	0.094
			Post-repetition	1.82	.60			
	Slow	20	Pre-repetition	1.56	.47	.27	2.046	0.427
			Post-repetition	1.83	.75			

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 August 2018 doi:10.20944/preprints201808.0522.v1

Table 3. *Cont.*

Emotion	Positive	60	Pre-repetition	1.99	.67	.29	3.258**	0.399
			Post-repetition	2.28	.80			
IV	Levels	*N*	Rating	Mean	*SD*	Mean rating change	*t*	*d*
Emotion	Negative	60	Pre-repetition	1.74	.65	.21	2.475*	0.286
			Post-repetition	1.95	.82			
	Neutral	60	Pre-repetition	1.45	.50	.17	2.199*	0.286
			Post-repetition	1.62	.71			
Language tonality	Cantonese (Tonal)	60	Pre-repetition	1.56	.56	.22	2.749**	0.330
			Post-repetition	1.78	.71			
	English (Non-tonal)	60	Pre-repetition	1.85	.66	.28	3.097**	0.384
			Post-repetition	2.13	.81			

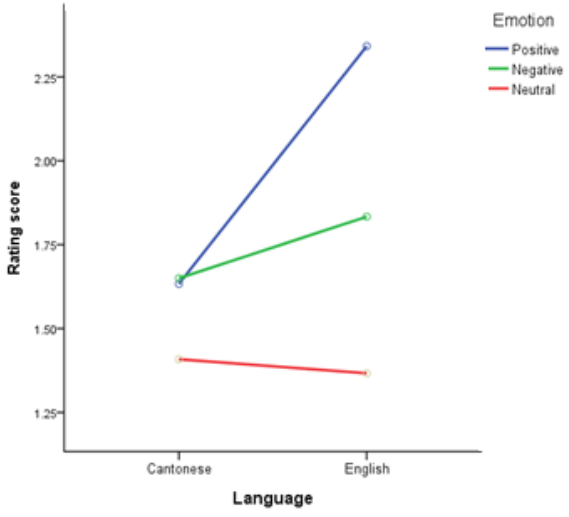
Note. * p-value < .05, ** p-value < .01, *** p-value < .001

3.2. Separate analysis of the ratings

In analysing each ratings, two 3 (pace) x 3 (emotion) x 2 (language tonality) ANOVAs were computed, one on the pre-repetition rating, another one on the post-repetition rating. See Table 4-6 for the effect of the independent variables on the two ratings.

3.2.1. Pre-repetition rating

There were significant main effects for emotion ($F(2, 56) = 45.936$, $p = .000$, $\eta_p^2 = .621$) and language tonality ($F(1, 57) = 14.792$, $p = .000$, $\eta_p^2 = .206$). Besides, the emotion x language tonality interaction was significant, $F(2, 56) = 26.360$, $p = .000$, $\eta_p^2 = .485$. As shown in Figure 5, the effect of emotion on the pre-repetition score differs depending on language tonality. Overall, neutral stimuli received the lowest score in the two languages. When the stimuli were provided in Cantonese, the scores for positive emotion and negative emotion were very similar (Positive emotion: $M = 1.63$, $SE = .09$; Negative emotion: $M = 1.65$, $SE = .09$). When the stimuli were given in English, there was a greater difference in the score among the three types of emotion.



Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 August 2018 doi:10.20944/preprints201808.0522.v1

Figure 5. The interaction of emotion and language of the pre-repetition rating.

3.2.2. Post-repetition rating

Different emotions and the two languages lead to difference in perception in the pre-repetition rating, and this pattern persisted to the post-repetition rating. The main effects for emotion ($F(2, 56) = 35.161$, $p = .000$, $\eta_p^2 = .557$) and language tonality ($F(1, 57) = 20.235$, $p = .000$, $\eta_p^2 = .262$) were significant. The interaction of emotion and language tonality has also reached the significance level, $F(2, 56) = 17.218$, $p = .000$, $\eta_p^2 = .381$ (see Figure 6).

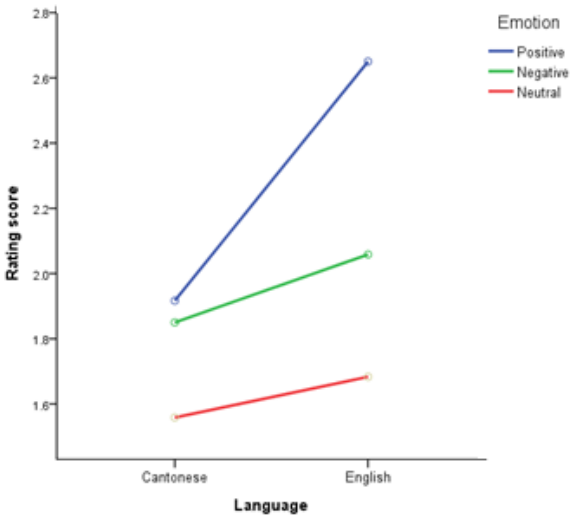


Figure 6. The interaction of emotion and language of the post-repetition rating.

Table 4. The Effect of Pace on the Pre-repetition Rating and the Post-repetition Rating.

Rating	Pace	*N*	Mean	*SE*	*F*	*p*	η_p^2
Pre-repetition	Fast	20	1.80	.12	1.134	.329	.038
	Normal	20	1.74	.12			
	Slow	20	1.39	.12			
Post-repetition	Fast	20	2.22	.15	2.229	.117	.073
	Normal	20	1.82	.15			
	Slow	20	1.83	.15			

Table 5. The Effect of Emotion on the Pre-repetition Rating and the Post-repetition Rating

Rating	Emotion	*N*	Mean	*SE*	*F*	*p*	η_p^2
Pre-repetition	Positive	60	1.99	.09	45.936	.000	.621
	Negative	60	1.74	.08			
	Neutral	60	1.39	.06			
Post-repetition	Positive	60	2.28	.10	34.268	.000	.555
	Negative	60	1.95	.10			
	Neutral	60	1.62	.09			

Table 6. The Effect of Language Tonality on the Pre-repetition Rating and the Post-repetition Rating

Rating	Language	*N*	Mean	*SE*	*F*	*p*	η_p^2
Pre-repetition	Cantonese	60	1.56	.07	14.792	.000	.206
	English	60	1.85	.09			
Post-repetition	Cantonese	60	1.78	.09	19.298	.000	.256
	English	60	2.13	.10			

4. Discussion

Briefly, this study found that the STS illusion was more obvious under the fast pace condition. The results also revealed that all three types of emotions (positive, negative, and neutral) and both two types of language tonality (tonal language and non-tonal language) could result in the STS illusion. Yet, not a particular type of emotion or language tonality could lead to the strongest experience of the illusion.

4.1. The Speech-to-song Illusion

Overall speaking, the post-repetition score was significantly higher than the pre-repetition score, indicating the role of repetition in changing our auditory perceptual experience. This is consistent with previous studies investigating the STS illusion [6,13,34,36].

Compared with previous studies, however, there are differences in the post-repetition score and the mean rating change. In the existing research, a contrast in scores is also observed when different stimuli were used. Using the stimulus recorded by Deutsch, she found that the mean rating score difference was 2.4 [6] and Vanden Bosch der Nederlanden et al. (2015) found a mean rating change larger than one, the score increased from 1.58 to 3.05 (i.e. the change of score was 1.47) [13]. Very differently, when self-selected stimuli were used, the mean rating change of the score of the stimuli were less than one [34,36].

The contrasting results may be due to the pitch variation of the recording. In Deutsch's recording, more pitch range are shown. Transforming her speech recording of the short excerpt (i.e. sometimes behave so strangely) to melody, it was D#4, D#4, C#4, B3, D#4, C#4, F3 [6]. Yet, even when the same phrase was spoken by other people, the pitch range was not that large. In the second experiment conducted by Deutsch et al. (2011), a group of participants repeated the phrase back after listening once to the stimulus, and the average melody was B3, B3, B3, A3, B3, A3, G3 [6].

Second, there is a cultural difference in responding to a Likert-scale. In the two previous studies, participants were all native English speakers, which are categorized as westerners. The current study recruited Hong Kong college student, which are categorized as Asians or Easterners. It was discovered that in answering questions on a Likert scale, Easterners tend to choose midpoints while Westerners tend to choose the extreme responses [40,41].

Third, participants in this current study may be influenced by their native language. All of them are native in Cantonese, which means they are native tonal language speakers. Their lower rating scores supported Jaisin et al. (2016)'s study that among native tonal language speakers, the experience of the STS illusion is reduced [34]. In tonal language, different pitch value of the syllable is encoded with different semantic meaning. Having been listening to tonal language since they were born, they inherently code pitch patterns as linguistic [34] and they are better at identifying pitch. In the study conducted by Mok and Zuo (2012), the native tonal language speakers performed more accurately than non-tonal language speakers in two AX discrimination tasks of Cantonese monosyllables and pure tones resynthesized from Cantonese tones [42]. Therefore, in the current study, the pitch they heard was easily identified, but it was not interpreted highly as song.

4.2. Pace

The findings that the STS illusion occurred in fast pace but not in slow pace in the current study is consistent with other research conducted on auditory illusion, specifically, the verbal

 doi:10.20944/preprints201808.0522.v1

transformation effect. A reduced effect was found when the presented target word was played slower [24,43].

The reason that people perceive speech presented in a fast pace as more song-like may possibly be related to our cognitive capacity and neural circuitry of pitch salience.

From Oakes and North (2006)'s experimental study, people had a lower recall rate of the advertisement information when fast-tempo background music was played [14]. This is because music in a faster pace distracted the person from information processing. Speaking of the STS illusion, Deutsch et al. (2011) made two hypotheses in regard to the occurrence of the illusory transformation [6]. First, there is the inhibition of the neural circuitry on pitch salience while people are listening to speech spoken in a normal pace, so they have full focus on the meaning of the speech. Then, when they hear several exact repetitions of the speech, there is an activation of their neural circuitry. This is supported by Tierney, Dick, Deutsch, and Sereno (2013)'s fMRI brain imaging research which discovered that presenting the phrases repeatedly activated brain regions like the anterior superior temporal gyrus for pitch processing [32]. In a later study, Margulis et al. (2015) explained the illusion using a terminology, namely, the semantic satiation effect [36]. It means that when people repeatedly listen to a same phrase, they would start to pay attention on other characteristics of the phrase (e.g. pitch), rather than keep on processing the meaning of the phrase. A faster pace which could distract people from processing information, together with the effect of the repetition of the phrase may be able to enhance the activation of the brain regions perceiving musical elements.

4.3. Emotion

This study revealed that all three types of emotions could result in the STS illusion. There was no difference among the three types in leading to the illusory change. This is unlike previous studies investigating on visual perception. It was found that people in positive mood process visual stimulus globally, thus, they are more likely to be affected by the context information in the Ebbinghaus illusion task; whereas in negative mood, as people process the stimulus locally, they could make better judgment in the task [25,26]. It may be that the effect of emotion on visual perception is not the same on auditory perception.

Furthermore, the difference may be due to the design of the experiment. Even though a pilot study was conducted to understand the emotionality of the sentences and the sentences selected has reached the selection criteria, having emotion as a within-subject variable would mean that participants were listening to sentences in all three types of emotions. As a consequence, the effect of emotion induced by the sentences would reduce.

4.4. Language Tonality of the Stimuli

Similar to the effect of emotion, both tonal language stimulus (i.e. Cantonese) and non-tonal language stimulus (i.e. English) could lead to a significant STS illusion. This confirmed previous studies that the illusion could be seen in both tonal language and non-tonal language stimulus [6,13,34,36]. The current research suggests that the effect of the language tonality of the stimulus on the STS illusion do not differ significantly among bilingual speakers native in tonal language with non-tonal language as secondary language.

4.5. Other Observations

There is an alternative explanation for not having significant main effect in emotion and language tonality in the STS illusion. From the separate analysis of ANOVAs on the pre-repetition score and the post-repetition score, the effect of emotion and language tonality can be seen from the pre-repetition score and the effect sustained in the post-repetition test. In fact, the song-like effect has started even without the repetition of the phrases.

In emotion, its effect was significant in both ratings. Positive emotion sentences received the highest score while negative emotion sentences received the lowest score. It is wondered why the

speech is heard as more musical when it is in a positive emotion. This could possibly be explained from the field of positive psychology. According to Fredrickson (2001), positive emotion broadens people by making them more open to different kinds of information [44]. Just as mentioned in the previous parts, they process visual stimulus globally by looking at both the target information and context information [25,26]. So, after participants have heard the full sentence in the first phase, the influence of positive emotion already come into effect.

In language tonality, the significant main effects in both pre-repetition score and post-repetition score indicates that non-tonal language stimuli had a significantly higher rating than the tonal language stimulus. Why is the speech heard as more musical when it is spoken in English? To Hong Kong people, the non-tonal language is their second language. Though they have been learning English since they were in kindergarten, it is quite common to see people speaking English in Hong Kong accent, where the influence of Cantonese like speaking in a flat tone and in a certain discrete pitch is prominent. Sewell and Chan (2010) also identified several consonantal features in Hong Kong English accent [45]. For example, 76 % of the Hong Kong people substitute the TH sound by [d], like saying 'dose' for 'those'. This may suggest that the second language people learned appear to sound more musical than their native language.

As all the conditions characterized by emotion or language tonality had an increase of score, there was no significant result found in the effect of emotion and language tonality on the STS illusion. It is concluded that emotion and language tonality could lead to a different perception at the very beginning, yet these effects did not change after the repetition of the phrases and were not strong enough to bring a significant illusory transformation.

4.6. Implications

This research adds new colors to the existing knowledge about auditory perception of music and speech. According to the findings, we could understand that simply presenting speech in positive emotion spoken in English to native Cantonese speakers sounds more musical to them. We also understand that a repetitive speech in a fast pace sounds more musical. It shows that speech and music are not in a dichotomy, it seems like a continuum. In some occasions, for example, in a repetitive speech, a speech spoken in a fast pace, or speaking in a higher pitch range such as the infant-directed speech, the 'speech' would appear to be musical and song-like. Applying this in daily life, some promotions or announcements in advertisements can be presented with repetition of phrases so that it would be more appealing by sounding more musical.

Furthermore, the current research facilitates the understanding of song production. We can see that songs in every genre typically have a chorus part. This could be explained by the present research that repetition makes sound to be more musical. Speech can turn into music in a very simple way. What makes it happen is by having repetitive phrases and to be played in a relatively fast pace. So, in the creation of songs, we should bear in mind the presence of repetitive phrases or words throughout the songs. For songs to excite people, fast tempo can be employed as people turn to be more aware of the musical elements like melody and rhythm. Although fast tempo was found to sound more musical in this study, we are not suggesting that all songs have to take up a fast tempo simply because they may sound more musical. In creation of songs, meaning and purposes are equally important as its musical virtue. For songs to bring out important messages or arise certain sentiment, slower pace could be more appropriate.

4.7. Limitations and Future Study

The current study has its own limitation. First, it would be better to define emotion as a between-subject variable so that participants would only listen to one type of the emotion and stay with the emotion evoked. Second, there was a lack of control on the time participants chose their ratings. Some participants clicked on the second rating after several repetitions had been played; some took a while to rate after the recording was played. It was not sure if they rated within 2300 ms, a time period for rating score in Deutsch et al. (2011)'s research [6]. Third, there was a problem in the playing of the

recordings. Google Form was used as an online experimental platform, which depends on a good Wi-Fi connection. In a few occasions, the Wi-Fi connection was poor, thus the playing of the recording was sometimes delayed or disrupted. This may affect how those affected participants perceive what they heard. In the future, it would be better to use psychological programs (e.g. E-Prime, PsychoPy) that does not rely on internet connection.

Although slightly higher scores in non-tonal stimuli (English stimuli) can be observed in this study, we are cautious not to make a firm conclusion on the effect of language tonality. We have the concern that the differences could be due to the effect of familiarity of the language as English is, after all, the second language of Hong Kong people. To differentiate the effect of language tonality and the effect of language familiarity, future study is needed to test bilingual speakers who can speak both one tonal language and one non-tonal language equally well.

Also, as mentioned in previous parts, the contrast of scores and the contrast in the effect of pace in the two experiments may possibly due to a larger pitch variation in Deutsch's stimulus. Meanwhile, Graber, Simchy-Gross, and Margulis (2017) stated that the stimuli that could lead to the STS illusion had music-like, isochronous beats [46]. And we can see from people playing any kind of drums that the sound does not have a large variation in pitch, yet it is still musical because of the regular beat pattern. Future study could explore if the pitch variation, the pace, or the rhythm of the sound is more vital to make sound more musical.

Moreover, brain imaging studies could be done to further investigate the STS illusion so as to understand more about how brain regions for language comprehension and pitch processing work, how these regions interact to perceive a sound, and whether there is more activation of the pitch processing areas when fast pace speech or fast pace music was played.

It would also be interesting to study people who has perfect pitch ability, congenital amusia, or aphasia to understand how they process sound and how they perceive the STS illusion. Additionally, the findings may be able to help people with congenital amusia. Congenital amusia is defined by having the difficulty in processing and recognizing pitch differences. People with amusia could understand language well and speak languages fluently but have difficulties in perceiving music [47,48]. The STS illusion connects the language domain and the music domain. Future study could investigate whether the stimulus that invoke the most STS illusion, such as the stimulus recorded by Deutsch, would activate the musical pitch processing when these people listen to the stimulus.

5. Conclusions

To conclude, this study shows that faster speech pace could result in stronger effect of the speech-to-song illusion. It was also found that neither emotion nor language tonality has a main effect on the speech-to-song illusion. In all, this is a pioneering study that explores the role of pace, emotion and language tonality in the speech-to-song illusion. Its findings enable us to understand that speech and song perception is not in a dichotomy but in a continuum.

6. Patents

Funding: This research received no external funding.

Acknowledgments: The principles and procedures used in the study concerning human research ethics were approved by the Human Research Ethics Committee of Hong Kong Shue Yan University. We would like to thank Cheryl Chan for her help in the preparation of audio stimuli.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Romer, F. *The Physiology of the Human Voice*. Kessinger Publishing: Montana, United States, 2010, ISBN 9781165072088.
2. Moore, B. C.; Patterson, R. D.; Winter, I. M.; Carlyon, R. P.; Gockel, H. E. *Basic Aspects of Hearing Physiology and Perception*; Springer: New York, United States, 2013, ISBN 9781493900183.

3. Stanley, D. *Your Voice: Applied Science of Vocal Art, Singing and Speaking*. Pitman Publishing Corporation: New York, United States, 1945.
4. Adams, T. M.; Fuller, D. B. The words have changed but the ideology remains the same: Misogynistic lyrics in rap music. *J Black Stud* **2006**, *36*, 938-957. doi:10.1177/0021934704274072
5. Uhlig, S.; Dimitriadis, T.; Hakvoort, L.; Scherder, E. Rap and singing are used by music therapists to enhance emotional self-regulation of youth: Results of a survey of music therapists in the Netherlands. *Arts Psychother* **2017**, *53*, 44-54. doi:10.1016/j.aip.2016.12.001
6. Deutsch, D.; Henthorn, T.; Lapidis, R. Illusory transformation from speech to song. *J Acoust Soc Am* **2011**, *129*, 2245-2252. doi:10.1121/1.3562174
7. The Chainsmokers take "#SELFIE" to No. 1 on dance chart. Available online: http://www.billboard.com/articles/columns/code/5944641/the-chainsmokers-take-selfie-to-no-1-on-dance-chart (accessed on 2 March 2017).
8. Sternberg, R. J.; Sternberg, K. *Cognitive Psychology*, 6th ed.; Cengage Learning: Belmont, United States, 2012, ISBN 9781133313915.
9. Terrell, S. H. Elements of Music and Speech: A Methodology to Incorporate the Elements of Music into Teaching Pronunciation to Speakers of English as a Second Language. Ph.D. Thesis, the University of Texas at Dallas, Richardson, May 2012.
10. Richards, J. C.; Schmidt, R. *Longman Dictionary of Language Teaching and Applied Linguistics*. Longman: New York, United States, 2002, ISBN 9781408204603.
11. Morrill, T. H.; Mcauley, J. D.; Dilley, L. C.; Hambrick, D. Z. Individual differences in the perception of melodic contours and pitch-accent timing in speech: Support for domain-generality of pitch processing. *J Exp Psychol* **2015**, *144*, 730-736, doi:10.1037/xge0000081
12. Patel, A. D.; Iversen, J. R.; Rosenberg, J. C. Comparing the rhythm and melody of speech and music: The case of British English and French. *J Acoust Soc Am* **2006**, *119*, 3034-3047, doi:10.1121/1.2179657
13. Vanden Bosch der Nederlanden, C. M.; Hannon, E. E.; Snyder, J. S. Everyday musical experience is sufficient to perceive the speech-to-song illusion. *J Exp Psychol* **2015**, *144*, 43-49, doi:10.1037/xge0000056
14. Kuribayashi, R.; Nittono, H. Speeding up the tempo of background sounds accelerates the pace of behaviour. *Psychol Music* **2015**, *43*, doi:10.1177/0305735614543216
15. Milliman, R. E. Using background music to affect the behavior of supermarket shoppers. *J Mark* **1982**, *46*, 86-91, doi:10.2307/1251706
16. Oakes, S.; North, A. C. The impact of background musical tempo and timbre congruity upon ad content recall and affective response. *Appl Cogn Psychol* **2006**, *20*, 505-520, doi:10.1002/acp.1199
17. Roballey, T.; McGreevy, C.; Rongo, R.; Schwantes, M.; Steger, P.; Wininger, M.; Gardner, E. The effect of music on eating behaviour. *Bull Psychon Soc* **1985**, *23*, 221-222.
18. Dilley, L. C.; Pitt, M. A. Altering context speech rate can cause words to appear or disappear. *Psychol Sci* **2010**, *21*, 1664-1670, doi:10.1177/0956797610384743
19. Dupoux, E.; Green, K. Perceptual adjustment to highly compressed speech: Effects of talker and rate changes. *J Exp Psychol Hum Percept Perform* **1997**, *23*, 914-927. doi:10.1037//0096-1523.23.3.914
20. Morrill, T.; Baese-Berk, M.; Heffner, C.; Dilley, L. Interactions between distal speech rate, linguistic knowledge, and speech environment. *Psychon Bull Rev* **2015**, *22*, 1451-1457, doi:10.3758/s13423-015-0820-9
21. Baese-berk, M. M.; Heffner, C. C.; Dilley, L. C.; Pitt, M. A.; Morrill, T. H.; McAuley, J. D. Long-term temporal tracking of speech rate affects spoken-word recognition. *Psychol Sci* **2014**, *25*, 1546-1553, doi: 10.1177/0956797614533705
22. Bosker, H. R. How our own speech rate influences our perception of others. *J Exp Psychol Learn Mem Cogn* **2017**, *43*, 1225-1238.
23. Warren, R. M.; Gregory, R. L. An auditory analogue of the visual reversible figure. *Am J Psychol* **1958**, *71*, 612-613. doi:10.2307/1420267
24. Castro, N. What Turns Speech into Song? Investigations of the Speech-to-Song Illusion. Ph.D. Thesis, The University of Kansas, Lawrence, 2014.
25. Avramova, Y. R.; Stapel, D. A.; Lerouge, D. Mood and contest-dependence: Positive mood increases and negative mood decreases the effects of contest on perception. *J Pers Soc Psychol* **2010**, *99*, 203-214, doi:10.1037/a0020216

26. Gasper, K.; Clore, G. L. Attending to the big picture: Mood and global versus local processing of visual information. *Psychol Sci* **2002**, *13*, 34-40. doi:10.1111/1467-9280.00406
27. Bergman, P.; Västfjäll, D.; Tajadura-Jimenez, A.; Asutay. Auditory-induced emotion mediates perceptual categorization of everyday sounds. Front Psychol 2016, 7, doi:10.3389/fpsyg.2016.01565
28. Asutay, E.; Västfjäll, D. Perception of loudness is influenced by emotion. *PLoS ONE* **2012**, *7*, doi:10.1371/journal.pone.0038660
29. Siegel, E. H.; Stefanucci, J. K. A little bit louder now: Negative affect increases perceived loudness. *Emotion* **2011**, *11*, 1006-1011, doi:10.1037/a0024590
30. Deutsch, D. The octave illusion revisited again. *J Exp Psychol Hum Percept Perform* **2004**, *30*, 355-364, doi:10.1037/0096-1523.30.2.355
31. Magne, C.; Schon, D.; Besson, M. Musician children detect pitch violations in both music and language better than nonmusician children: Behavioural and electrophysiological approaches. *J Cogn Neurosci* **2006**, *18*, 199-211, doi:10.1162/089892906775783660
32. Tierney, A.; Dick, F.; Deutsch, D.; Sereno, M. Speech versus song: Multiple pitch-sensitive areas revealed by a naturally occurring musical illusion. *Cereb. Cortex* **2013**, *23*, 249-254, doi:10.1093/cercor/bhs003
33. Francis, A. L.; Ciocca, V.; Ma, L.; Fenn, K. Perceptual learning ofCantonese lexical tones by tone and non-tone language speakers. *J Phon* **2008**, *36*, 268-294. doi:10.1016/j.wocn.2007.06.005
34. Jaisin, K.; Suphanchaimat, R.; Figueroa Candia, M. A.; Warren, J. D. The speech-to-song illusion is reduced in speakers of tonal (vs. non-tonal) languages. *Front Psychol* **2016**, *7*, 1-9, doi:10.3389/fpsyg.2016.00662
35. Tong, X.; Lee, S. M. K.; Lee, M. M. L.; Burnham, D. A tale of two features: Perception of Cantonese lexical tone and English lexical stress in Cantonese-English bilinguals. *PLoS ONE* **2015**, *10*, doi:10.1371/journal.pone.0142896
36. Margulis, E. H.; Simchy-Gross, R.; Black, J. L. Pronunciation difficulty, temporal regularity, and the speech-to-song illusion. *Front Psychol* **2015**, *6*, 1-7, doi:10.3389/fpsyg.2015.00048
37. Audacity(R): Free Audio Editor and Recorder (Version 2.1.3). Available online: http://www.audacityteam.org/download/ (accessed on 10 July 2017)
38. Praat: Doing phonetics by computer (Version 6.0.28). Available online: http://www.praat.org (accessed on 2 April 2017)
39. Research Randomizer. Available online: https://www.randomizer.org/ (accessed on 17 September 17 2017)
40. Lee, J. W.; Jones, P. S.; Minetama, Y.; Zhang, X. E. Cultural differences in responses to a likert scale. *Res Nurs Health* **2002**, *25*, 295-306, doi:10.1002/nur.10041
41. Wang, R.; Hempton, B.; Dugan, J. P.; Komives, S. R. Cultural differences: Why do asians avoid extreme responses? *Surv Pract* **2008**, *1*, doi:10.29115/sp-2008-0011
42. Mok, P. K.; Zuo, D. The separation between music and speech: Evidence from the perception of Cantonese tones. *J Acoust Soc Am* **2012**, *132*, 2711-2720, doi:10.1121/1.4747010
43. Synder, K. A.; Calef. R. S.; Choban, M. C. Effects of word repetition and presentation rate on the frequency of verbal transformations: Support for habituation. *Bull Psychon Soc* **1993**, *31* 91-93, doi:10.3758/bf03334148
44. Fredrickson, B. L. The role of positive emotions in positive psychology: The broaden-and-build theory of positive emotions. *Am Psychol* **2001**, *56*, 218-226.
45. Sewell, A.; Chan, J. Patterns of variation in the consonantal phonology of Hong Kong English. *EWW* **2010**, *31*, 138-161, doi:10.1075/eww.31.2.02sew
46. Graber, E.; Simchy-Gross, R.; Margulis, E. H. Musical and linguistic listening modes in the speech-to-song illusion bias timing perception and absolute pitch memory. *J Acoust Soc Am* **2017**, *142*, 3593-3602, doi:10.1121/1.5016806
47. Ayotte, J.; Peretz, I.; Hyde, K. Congenital amusia: A group study of adults afflicted with a music-specific disorder. *Brain* **2002**, *123*, 238-251.
48. Stewart, L. Congenital amusia. *Curr Biol* **2006**, *16*, 904-906.