Title: Repetition enhances the musicality of speech and tone stimuli to similar degrees

Running Head: Repetition enhances musicality of speech and tones

Authors: Adam Tierney[1], Aniruddh D. Patel[2,3], Mara Breen[4]

Affiliations

[1]Department of Psychological Sciences, Birkbeck, University of London, London, UK

[2]Department of Psychology, Tufts University, Medford, MA

[2]Canadian Institute for Advanced Research (CIFAR), Azrieli Program in Brain, Mind, and Consciousness, Toronto, CA

[4]Psychology and Education Department, Mount Holyoke College, South Hadley, MA

IN PRESS AT MUSIC PERCEPTION

Abstract

Certain spoken phrases, when removed from context and repeated, begin to sound as if they were sung. Prior work has uncovered several acoustic factors which determine whether a phrase sounds sung after repetition. However, the reason why repetition is necessary for song to be perceived in speech is unclear. One possibility is that by default pitch is not a salient attribute of speech in non-tonal languages, as spectral information is more vital for determining meaning. However, repetition may satiate lexical processing, increasing pitch salience. A second possibility is that it takes time to establish the precise pitch perception necessary for assigning each syllable a musical scale degree. Here we tested these hypotheses by asking participants to rate the musicality of spoken phrases and complex tones with matching pitch contours after each of eight repetitions. Although musicality ratings were overall higher for the tone stimuli, both the speech and complex tone stimuli increased in musicality to a similar degree with repetition. Thus, although the rapid spectral variation of speech may inhibit pitch salience, this inhibition does not decrease with repetition. Instead, repetition may be necessary for the perception of song in speech because the perception of exact pitch intervals takes time.

Keywords: speech, singing, pitch, perception, language

Introduction

Speech and music are generally studied as if they were distinct categories. For example, there have been attempts to construct automated methods for distinguishing speech and music based on acoustic characteristics (Schluter & Sonnleitner, 2012). However, certain spoken phrases, if removed from context and repeated, can be perceived as song, suggesting instead that speech and music are acoustically overlapping categories. The first demonstration of this phenomenon described a striking single example (Deutsch, Lapidis, & Henthorn, 2011), showing that exact repetition (i.e., looping of a short spoken phrase) was necessary for the transformation to take place. Participants were additionally asked to repeat back what they heard either after a single presentation or after several repetitions, and the increase in song perception was linked to more accurate repetition of the underlying pitch contour.

This phenomenon demonstrates that music perception is a listening mode which can be applied to verbal stimuli which were not originally intended to be heard as music. Several follow-up studies on this phenomenon have been published in recent years focusing on which stimulus characteristics are linked to stronger song percepts or more rapid transformations. Tierney, Dick, Deutsch, & Sereno (2013), for example, showed that the phenomenon was replicable in a larger sample of illusion stimuli, and that they could be matched to a set of control stimuli which do not transform. Der Nederlanden, Hannon, & Snyder (2015a) confirmed this distinction between illusion and control stimuli in a group of non-musician participants, and demonstrated that the illusion affected the accuracy of pitch discrimination (Der Nederlanden, Hannon, & Snyder 2015b). Falk et al. (2014) demonstrated that the speed of the speech-to-song transformation could be modulated by manipulating the flatness of pitch contours, the presence of a scalar interval, and rhythmic regularity. Finally, Margulis, Simchy-Gross, & Black (2015) found that passages from less pronounceable languages were perceived as more musical after repetition.

This body of work confirms that spoken stimuli can be perceived as song and identifies a range of acoustic, linguistic, and musical characteristics that influence the strength of this musical percept. This suggests that music perception is a listening mode that can be applied to a wide range of stimuli, including speech, so long as certain preconditions are present. However, it remains unclear why stimulus repetition is necessary for song perception to take place. That is, if the necessary preconditions are present, why does the stimulus not sound song-like immediately? One possibility, the *spectral salience hypothesis*, is that to comprehend speech listeners need to direct attention to spectral information in order to follow the rapid spectro-temporal changes which convey different phonetic categories. Thus, spectral information tends to capture attention, causing the salience of pitch information to be initially low. According to this account (Deutsch et al., 2011; Tierney et al., 2013), stimulus repetition leads to satiation of lexical nodes (Smith & Klein, 1990), causing the salience of pitch information to rise. This would explain why less pronounceable languages are perceived as more musical after repetition (Margulis et al., 2015): they are captured less by speech perception mechanisms, thus increasing pitch salience. This account is also supported by work showing that pitch perception is less accurate for stimuli that include greater spectral shape variation (Allen & Oxenham, 2014; Caruso & Balaban, 2014; Warrier and Zatorre, 2002), indicating a trade-off between spectral and pitch perception.

Another possibility, the *melodic structure hypothesis*, is that repetition is necessary for song perception to take place because melodic structure takes time to extract from the stimuli. In order to perceive a stimulus as song, listeners must decide which musical scale best fits the sequence of pitches, then assign each syllable a particular degree on this scale. This requires participants to rapidly encode into short-term memory a set of exact intervals between pitches so that these intervals can be compared to a number of different scale templates. However, if simple tone sequences are presented only once, listeners generally retain only the melodic contour (Dowling, 1978), and further repetitions are necessary to enable identification of exact intervals (Deutsch, 1979). This account is supported by work showing that random tone sequences are rated as more musical and more enjoyable after repetition (Margulis, 2013a; Margulis & Simchy-Gross, 2016) and work showing that explicit memory for novel melodies is relatively poor after a single presentation (Bartlett, Halpern, & Dowling, 1995).

Here we tested these hypotheses by synthesizing complex tones which followed the pitch contour of Illusion and Control stimuli drawn from the corpus of Tierney et al. (2013). These stimuli, therefore, contained the same pitch information as the original stimuli but no spectral variation. We then asked two groups of participants to rate the musicality of the original speech stimuli and the complex tone stimuli, respectively, after each of eight repetitions. If spectral salience is entirely responsible for the increase in musicality with repetition, then the complex tone Illusion stimuli should sound highly musical after a single presentation but not increase in musicality with repetition, and the difference in musicality between Illusion and Control stimuli should be initially large and not increase with repetition. On the other hand, if melodic structure is entirely responsible for the repetition effect, then the speech and complex tone stimuli should increase in musicality to the same degree with repetition. Finally, if both spectral salience and melodic structure are responsible for the repetition effect, then musicality judgments of the speech and complex tone stimuli should both increase with repetition, but the repetition effect should be larger for the speech stimuli.

Methods

Experiment design

Stimulus type (speech versus complex tone) was manipulated using a between-subjects design. Although a within-subjects design would provide more statistical power, it is vulnerable to effects of prior exposure to a particular stimulus. For example, having previously heard the complex tone version of a stimulus could cue listeners in to the underlying pitch contour, thereby diminishing the magnitude of the increase in musicality with repetition upon exposure to the matching speech stimulus.

Participants

32 participants (24 female) completed the speech stimulus experiment. Their mean age was 29.6 (standard deviation 7.1) years, and they had an average of 2.7 (3.2) years of musical training. 32 participants (23 female) completed the complex tone stimulus experiment. Their mean age was 31.2 (7.9) years, and they had an average of 3.8 (7.4) years of musical training. Thus the groups did not differ significantly in age ($t = 0.89$, $p = 0.38$) or musical training ($t = 0.72$, $p = 0.47$). Participants were compensated with either class credit or a payment of £5. All experimental procedures were approved by the Ethics Committee of the Department of Psychological Sciences at Birkbeck, University of London. Informed consent was obtained from all participants.

Stimuli

Speech stimuli consisted of 48 spoken phrases from audiobooks, obtained with permission from librivox.org and audiobooksforfree.com. It could be inferred from the context in which the phrases were produced that they were all originally intended to be heard as speech. This stimulus set was constructed via exhaustive search of audiobook sources for stimuli which either sound strongly musical (“illusion” stimuli) or not musical whatsoever (“control” stimuli) after repetition. Prior research using this stimulus set (Tierney et al., 2013) has confirmed that participants more often report a transformation from speech to song after repetition for the illusion stimuli, as compared to the control stimuli. The illusion and control stimulus sets are matched for speakers and number of syllables. More details about this stimulus set can be found in Tierney et al. (2013).

Complex tone stimuli were constructed via modification of the speech stimuli using the following procedure. First, the pitch contour of each phrase was extracted using the autocorrelation method with default settings in Praat (Boersma & Weenink, 2017). The resulting contour was then manually corrected to remove spurious octave jumps. Six-harmonic complex gliding tones were then constructed via custom Matlab scripts with a fundamental frequency equal to the phrase's pitch contour, and with equal amplitude across the six harmonics. Portions of the speech stimuli for which Praat did not extract a pitch contour were replaced with silence. A 10-ms cosine ramp was applied at each boundary between tone and silence to eliminate transients. See Figure 1 for an example of waveforms and spectrograms of the speech and complex tone versions of an example stimulus. These audio examples are also available for download in Supplementary Information.

[Insert Figure 1 about here]

Procedure

The experiment was conducted using HTML5. The participant was seated in front of a computer screen featuring the instructions "Listen to this passage and rate how musical it sounds, using the scale below," and a button labelled "Start trial". The instructions remained onscreen for the duration of the experiment. After the participant pressed the start trial button, one of the 48 stimuli was presented eight times. Stimulus order was randomized for each participant. After each presentation, a set of ten boxes containing the numerals 1 through 10 was simultaneously displayed on screen, along with the labels "non-musical" and "musical" aligned with the lowest- numbered and highest-numbered boxes, respectively. (This procedure differs slightly from that of Deutsch et al. (2011), who asked participants to rate the stimulus on a 1 to 5 scale. Here, a 1 to 10 scale was chosen to allow participants a slightly greater degree of granularity when making musicality judgments.) Clicking on one of these boxes caused the program to immediately advance to the next repetition. If the participant did not click on a box within two seconds, the boxes disappeared, and the next repetition was begun. This two-second timeout was imposed to ensure that each participant was exposed to a rapid series of repetitions of each stimulus. This procedure resulted in occasional missing data points for a particular repetition of a given stimulus. These missing data points (less than 1% of the total dataset) were replaced with the mean of the nearest prior and subsequent rating.

Results

Musicality ratings following each repetition are displayed in Figure 2. First, means and standard deviations were calculated across items. For the speech stimuli, musicality ratings of the illusion tokens increased from 3.98 (0.87) to 5.10 (1.01), while ratings of the control tokens increased from 3.15 (0.37) to 3.39 (0.44). For the complex tone stimuli, musicality ratings of the illusion tokens increased from 5.31 (0.62) to 6.67 (0.72), while ratings of the control tokens increased from 2.83 (0.42) to 3.41 (0.57). Thus, for both speech and complex tone stimuli, musicality ratings increased with repetition and were higher for illusion stimuli than for control stimuli.

[Insert Figure 2 about here]

We used linear mixed-effects regression to investigate whether the magnitude of the increase in musicality with repetition and the difference in musicality between illusion and control stimuli differed for speech and complex tone stimuli. Fixed effects were repetition (one through eight), stimulus set (illusion versus control), and experiment (speech versus complex tone). Random effects included intercepts for subjects and items, as well as repetition-by-subject and repetition-by-item slopes. Model parameters are listed in Table 1. P-values were calculated using the Wald test.

[Insert Table 1 about here]

There was a main effect of repetition ($B = 0.27$, $p < 0.01$), indicating that musicality ratings increased with repetition, and a main effect of stimulus set ($B = 0.87$, $p < 0.01$), indicating that illusion stimuli were rated as more musical than control stimuli. There was also an interaction between repetition and stimulus set ($B = -0.14$, $p < 0.01$), indicating that the increase in musicality with repetition was greater for the illusion than for the control stimuli. There was a main effect of experiment ($B = 3.2$, $p < 0.05$), indicating that musicality ratings were greater for the complex tone stimuli than for the speech stimuli, and an interaction between experiment and stimulus set (B = -1.77, p < 0.01), indicating that the rating difference between illusion and control stimuli was greater for the complex tone stimuli. However, and crucially, there was not an interaction between repetition and experiment (B = 0, p = 0.28). This indicates that there was no difference between the speech and complex tone stimuli in the size of the increase in musicality with repetition. There was also no three-way interaction between repetition, stimulus set, and experiment (B = 0.02, p = 0.29), indicating that the greater increase in musicality with repetition for the illusion stimuli compared to the control stimuli did not differ between the speech and complex tone stimuli.

There were large differences across stimuli in the extent to which they were rated as musical after repetition. For the speech stimuli, for example, musicality ratings after the eighth repetition ranged from 2.69 to 7.53. To investigate whether the cues to musicality were similar between the speech and complex tone stimulus sets, we first computed averaged musicality ratings after the eighth repetition across subjects for each stimulus. We then measured the relationship between musicality ratings of the speech stimuli and their matching complex tone stimuli using Spearman's correlations. Speech and complex tone ratings were correlated (rho = 0.73, p < 0.01), indicating that the speech stimuli which sounded highly musical after repetition also tended to sound highly musical even when presented in complex tone form. A scatterplot displaying the relationship between ratings of speech and complex tone stimuli can be found in Figure 3.

[Insert Figure 3 about here]

Discussion

We found that listeners judged speech stimuli as more musical after repetition, and that this increase in musicality was greater for a set of pre-defined "illusion" stimuli compared to "control" stimuli. This finding replicates the basic speech-to-song illusion effect reported in Tierney et al. (2013). However, we found that the increase in musicality with repetition and the difference in the size of the repetition effect between illusion and control stimuli was present to the same degree for complex tone sequences with the same pitch contour as the original stimuli.

These results indicate that spectral salience cannot be the primary explanation for why repetition is necessary for speech stimuli to be perceived as song, since the same pattern of transformation is perceived for spectrally simple versions of the same stimuli. Instead, our findings suggest that stimulus repetition makes possible extraction of the pitch information necessary for building a mental model of scale structure. In order for the pitch contours underlying syllables to be assigned to scale degrees, two main processing steps must be completed. First, each syllable must be assigned a single steady pitch, despite the existence of pitch variability within syllables. Second, the exact intervals between syllables must be calculated, so that the scale structure best fitting the sequence of pitches can be calculated. Future work could investigate which of these two steps is responsible for the repetition effect by investigating the size of the repetition effect for gliding-tone versus static-tone stimuli. It is important to note, however, that our results are not exclusive of other explanations for the impact of repetition on musicality. Other factors, including the facilitation of entrainment and imagined imitation (Margulis, 2013b), could contribute to the increase in musicality with repetition. Nevertheless, what can be decisively concluded from our findings is that the presence of speech information is not the driving factor underlying the repetition effect.

Our results indicate that variation in spectral shape can inhibit perception of the musicality of speech: complex tone stimuli were rated as more musical *overall*, both after a single presentation and after repetition. These results are in line with prior demonstrations that the presence of spectral shape variation can interfere with pitch perception (Allen & Oxenham, 2014; Caruso & Balaban, 2014; Warrier & Zatorre, 2002). However, our results suggest that the extent of this spectral interference does not decrease with repetition. This account helps explain the finding of Margulis et al. (2015) that less pronounceable languages sounded more musical than more pronounceable languages both before *and* after repetition: more pronounceable languages may have increased spectral salience, and the consequences of this up-regulated processing of speech information may not decrease with repetition.

The strength of the relationship we find between individual differences in the musicality of the original stimuli and the musicality of the complex tone versions of the same stimuli suggests that linguistic features (such as phonological neighbourhood, syntactic complexity, stress patterns, etc.) cannot be the primary factor differentiating stimuli which do transform and stimuli that do not, at least in this particular stimulus set. Indeed, there is sufficient information present in the signal to differentiate between musical and non-musical speech even when all spectral and linguistic content as well as much of the rhythmic information is filtered out. This suggests that pitch-based characteristics such as the flatness of pitch contours within syllables (Lindblom & Sungberg 2007; Schluter & Sonnleitner 2012) and the presence of musical intervals (Falk et al. 2014) may be the most important factor driving whether a given stimulus transforms from speech to song.

References

Allen, E. J., & Oxenham, A. J. (2014). Symmetric interactions and interference between pitch and timbre. *The Journal of the Acoustical Society of America*, 135(3), 1371-1379.

Bartlett, J., Halpern, A., & Dowling, J. (1995). Recognition of familiar and unfamiliar melodies in normal aging and Alzheimer's disease. *Memory & Cognition*, 23, 531-546.

Boersma, P., & Weenink, D. (2017). Praat: doing phonetics by computer [Computer program]. Version 6.0.22, retrieved from http://www.praat.org/

Caruso, V. C., & Balaban, E. (2014). Pitch and timbre interfere when both are parametrically varied. *PloS one*, 9(1), e87065.

Der Nederlanden, C., Hannon, E., & Snyder, J. (2015a). Everyday musical experience is sufficient to perceive the speech-to-song illusion. *Journal of Experimental Psychology: General,* 2, e43-e49.

Der Nederlanden, C., Hannon, E., & Snyder, J. (2015b). Finding the music of speech: musical knowledge influences pitch processing in speech. *Cognition,* 143, 135-140.

Deutsch, D. (1979). Octave generalization and the consolidation of melodic information. *Canadian Journal of Psychology,* 33, 201-205.

Deutsch, D., Henthorn, T., & Lapidis, R. (2011). Illusory transformation from speech to song. *JASA,* 129, 2245-2252.

Dowling, J. (1978). Scale and contour: two components of a theory of memory for melodies. *Psychological Review,* 85, 341-354.

Falk, S., Rathcke, T., & Dalla Bella, S. (2014). When Speech Sounds Like Music. *Journal of Experimental Psychology: Human Perception and Performance,* 40, 1491-1506.

Lindblom, B., & Sundberg, J. (2007). The human voice in speech and singing. In *Springer Handbook of Acoustics* (pp. 669-712). Springer New York.

Margulis, E. (2013a). Aesthetic responses to repetition in unfamiliar music. *Empirical Studies of the Arts,* 31, 45-57.

Margulis, E. (2013b). *On Repeat: How Music Plays the Mind*. New York, NY: Oxford University Press.

Margulis, E., Simchy-Gross, R., & Black, J. (2015). Pronunciation difficulty, temporal regularity, and the speech-to-song illusion. *Frontiers in Psychology,* 6, 48.

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

Schluter, J., & Sonnleitner, R. (2012). Unsupervised feature learning for speech and music detection in radio broadcasts. In *Proceedings of the 15th International Conference on Digital Audio Effects*.

Smith, L., & Klein, R. (1990). Evidence for semantic satiation: repeating a category slows subsequent semantic processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16, 852-861.

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

Warrier, C., & Zatorre, R. (2002). Influence of tonal context and timbral variation on perception of pitch. *Perception and Psychophysics,* 64, 198-207.

Author Note

Correspondence concerning this article should be addressed to Adam Tierney, Department of Psychological Sciences, Birkbeck, University of London, Malet Street, London, WC1E 7HX. Email: a.tierney@bbk.ac.uk.

	B	*std. Error*	*p-value*
Fixed Parts			
(Intercept)	1.72	0.63	**<0.01**
Repetition	0.27	0.07	**<0.01**
Stimulus set	0.87	0.23	**<0.01**
Experiment	3.2	0.38	**<0.05**
Rep:StimSet	-0.14	0.04	**<0.01**
StimSet:Expt	-1.77	0.11	**<0.01**
Rep:Expt	0	0.05	0.28
Rep:StimSet:Expt	0.02	0.02	0.29
Random Parts			
N_{Item}			96
$N_{Subject}$			64
Observations			24576

Table 1. Model parameters for linear mixed effects models comparing effects of Repetition and Stimulus Set for each Experiment. P-values were computed using the Wald test.

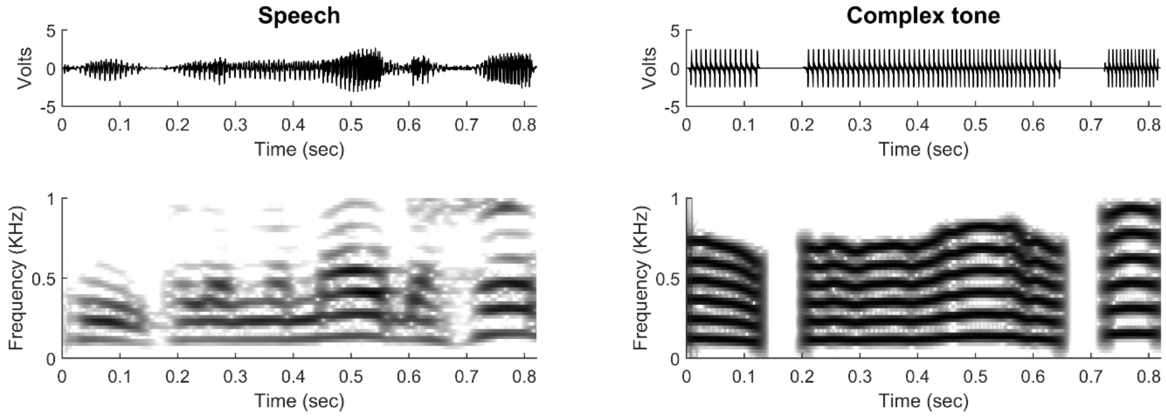

Figure 1. Waveform (top) and spectrogram (bottom) of an example stimulus, illustrating the difference between the speech (left) and complex tone (right) manipulations. Spectrograms were constructed using a 1024-point Hanning window (sample rate 22050 Hz) with an overlap of 958 time points, and were clipped at 40 dB below maximum value.

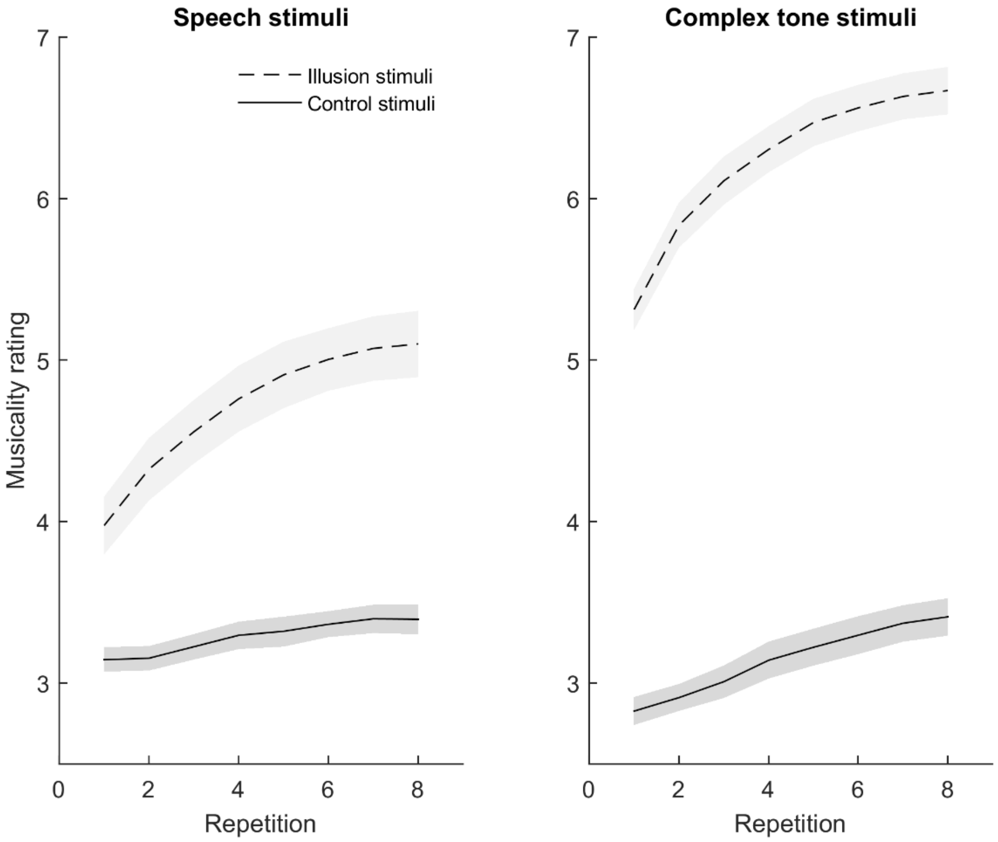

Figure 2. Increase in musicality with repetition for speech (left) and complex tone (right) stimuli across illusion (dotted line) and control (solid line) stimulus sets. The shaded regions indicate standard error of the mean.

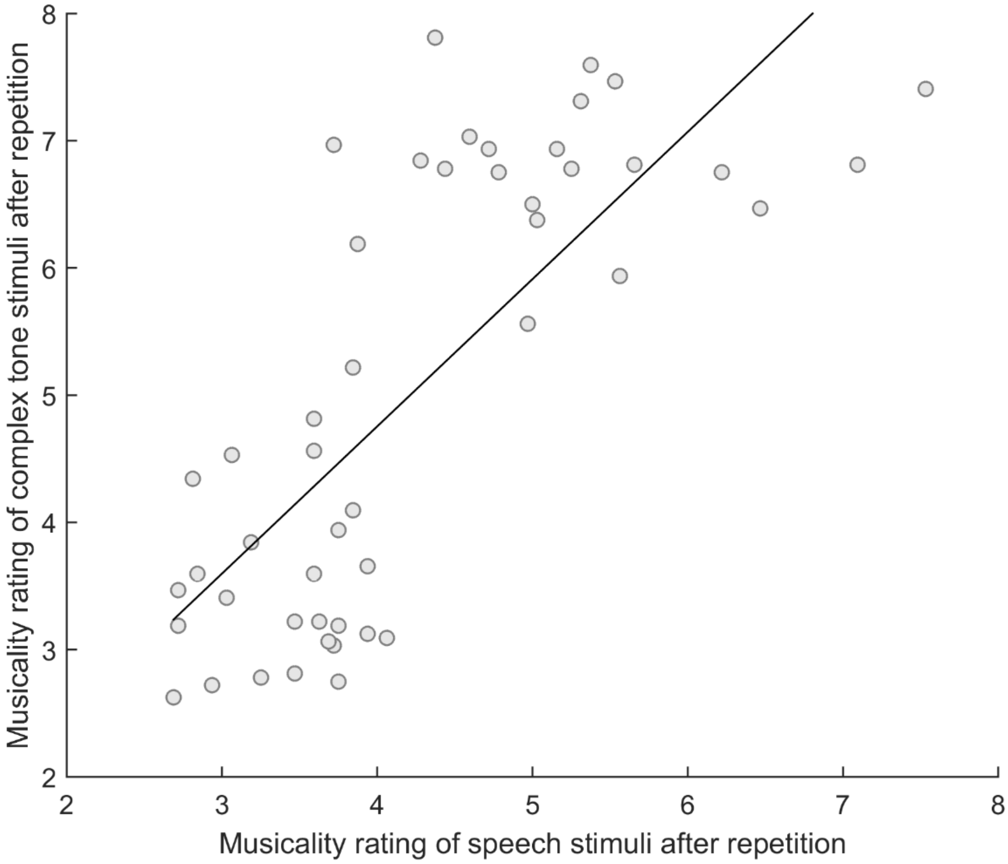

Figure 3. Scatterplot displaying the relationship between musicality ratings of matched speech and complex tone stimuli after the eight repetitions. Musicality ratings across the two stimulus types were positively correlated (rho = 0.73, p < 0.01).