Journal of Experimental Psychology:
Human Perception and Performance
2014, Vol. 40, No. 4, 1491–1506

© 2014 American Psychological Association
0096-1523/14/$12.00 http://dx.doi.org/10.1037/a0036858

When Speech Sounds Like Music

Simone Falk
Ludwig-Maximilians-University and Aix-Marseille Université

Tamara Rathcke
University of Glasgow and University of Kent

Simone Dalla Bella
University of Montpellier-1, Institut Universitaire de France, and University of Finance and Management in Warsaw

Repetition can boost memory and perception. However, repeating the same stimulus several times in immediate succession also induces intriguing perceptual transformations and illusions. Here, we investigate the *Speech to Song Transformation* (S2ST), a massed repetition effect in the auditory modality, which crosses the boundaries between language and music. In the S2ST, a phrase repeated several times shifts to being heard as sung. To better understand this unique cross-domain transformation, we examined the perceptual determinants of the S2ST, in particular the role of acoustics. In 2 Experiments, the effects of 2 pitch properties and 3 rhythmic properties on the probability and speed of occurrence of the transformation were examined. Results showed that both pitch and rhythmic properties are key features fostering the transformation. However, some properties proved to be more conducive to the S2ST than others. Stable tonal targets that allowed for the perception of a musical melody led more often and quickly to the S2ST than scalar intervals. Recurring durational contrasts arising from segmental grouping favoring a metrical interpretation of the stimulus also facilitated the S2ST. This was, however, not the case for a regular beat structure within and across repetitions. In addition, individual perceptual abilities allowed to predict the likelihood of the S2ST. Overall, the study demonstrated that repetition enables listeners to reinterpret specific prosodic features of spoken utterances in terms of musical structures. The findings underline a tight link between language and music, but they also reveal important differences in communicative functions of prosodic structure in the 2 domains.

Keywords: massed repetition, music perception, speech perception, transformation

"Repetition is the mother of knowledge," says a Latin proverb. A number of well-studied effects of repetition are in keeping with this ancient wisdom. In priming, a single repetition of a visual, tactile, or speech stimulus can considerably reduce the time and neural activation needed for its processing (see Horner & Henson, 2008, for a review; Gillmeister, Catmur, Liepelt, Brass, & Heyes, 2008; Kujala, Vartiainen, Laaksonen, & Salmelin, 2012; Zago, Fenske, Aminoff, & Bar, 2005). Multiple repetitions significantly enhance memory and learning provided that stimuli are presented occasionally and with several intervening items (i.e., the so-called spacing effect; see Delaney & Verkoeijen, 2009; Raaijmakers, 2003; Xue et al., 2011). Interestingly, these benefits are challenged or even reversed during massed repetition, that is, when a stimulus is repeated several times in immediate succession. That type of repetition is not only less effective for learning (e.g., Delaney & Verkoeijen, 2009), but it may also induce perceptual distortions and transformations. In the verbal domain, repetition of vowel sequences leads to the perception of two voice streams uttering illusory syllables (Warren, 2008; Warren, Bashford, & Gardner, 1990; Warren, Healy, & Chalikia, 1996). If a word is repeated for longer than one minute, priming is inhibited, and listeners lose the sense of the meaning of the word as a result of semantic satiation (e.g., Cattaneo, Devlin, Vecchi, & Silvanto, 2009; Kounios, Kotz, & Holcomb, 2000; Pilotti & Khurshid, 2004; Smith, 1984). Additionally, verbal transformations may occur and listeners start to hear words that are actually not present in the acoustic signal (e.g., *face, space, paste* instead of *pace*, MacKay, Wulf, Yin, & Abrams, 1993; Warren, 1961, 1968; Warren & Gregory, 1958; for nonverbal stimuli, see Kaminska & Mayer, 2002). In sum, massed repetition prompts listeners to perceive something new and factually not present in the signal or to radically change the initial interpretation of the signal.

This article was published Online First June 9, 2014.

Simone Falk, Institute of German Philology, Ludwig-Maximilians-University, Munich, Germany and Laboratoire Parole et Langage, CNRS, UMR 7309, Aix-Marseille Université, Aix-en-Provence, France; Tamara Rathcke, Department of English Language, University of Glasgow, Glasgow, United Kingdom and Department of English Language and Linguistics, University of Kent, Kent, United Kingdom; Simone Dalla Bella, Movement to Health Laboratory, EuroMov, University of Montpellier-1, Montpellier, France, Institut Universitaire de France, and Department of Cognitive Psychology, University of Finance and Management in Warsaw, Warsaw, Poland.

We thank the Institute of Phonetics and Speech Processing at the LMU Munich for insightful discussions and use of lab facilities and Sam Miller for statistical advice. We are grateful to Diana Deutsch, Salvador Soto-Faraco, and two anonymous reviewers who offered valuable comments on a previous version of the manuscript. This work was supported by a research grant of the LMU Munich to the first author and a travel grant of the Royal Society of Edinburgh to the second author.

Correspondence concerning this article should be addressed to Simone Falk, Institut für deutsche Philologie, Ludwig-Maximilians-Universität, Schellingstr.3, 80799 München, Germany. E-mail: simone.falk@germanistik.uni-muenchen.de

In this study, we investigate a massed repetition effect that induces a perceptual transformation from speech to song (Deutsch, 2003; Deutsch, Henthorn, & Lapidis, 2011; Tierney, Dick, Deutsch, & Sereno, 2012). We will refer to this phenomenon as the "*Speech to Song Transformation*" (henceforth S2ST). In general, listeners have no difficulties in telling the difference between spoken and sung sequences as they pertain to two forms of communication (i.e., language and music, respectively) that are functionally and structurally distinct. Yet, when a spoken sentence is repeated over and over again, listeners often report a perceptual transformation into song (Deutsch et al., 2011). It is well-known that listeners can perceive certain nonspeech signals as speech (e.g., sinewave speech, phoneme restoration; Groppe et al., 2010; Hillenbrand, Clark, & Baer, 2011; Remez, Rubin, Pisoni, & Carrell, 1981; Samuel, 1981; Warren, 1970). However, the S2ST is the only transformation from a speech percept to a musical percept that has, to our knowledge, been described.

In the original experimental study on the S2ST (Deutsch et al., 2011), musically trained participants listened 10 times to the English phrase "sometimes behave so strangely" which was taken from a longer utterance and repeated 10 times. After each repetition, the listeners were asked to judge on a 5-point scale whether the phrase sounded more like speech or more like song. Participants rated the phrase as spoken at the beginning and as sung at the end of the trial, in spite of the fact that the acoustics of the stimulus remained unchanged across repetitions. In fact, it was crucial for the occurrence of the S2ST that all acoustic features of the stimulus were held constant. Even a slight transposition of pitch (e.g., moving the pitch contour of the phrase up or down by 11/3 or 2/3 semitones) or a jumbled ordering of syllables within the stimulus during repetition were sufficient to hinder the transformation. Furthermore, choral singers were able to sing back the phrase after the repetitions, approximating a well-formed melody (Deutsch et al., 2011). Independent raters judged those productions as sung whereas productions following a single exposure to the test phrase were judged as spoken. That speech stimuli were indeed perceived as sung as a consequence of repetition was confirmed in a recent neuroimaging study (Tierney et al., 2012). Stimuli resulting in a S2ST elicited the same kind of heightened activation in several cortical areas (e.g., the anterior superior temporal gyrus, STG, and the midposterior STG) that has previously been found during song perception and production (e.g., Callan et al., 2006; Schön et al., 2010).

These results provide compelling evidence that repetition can induce a perceptual transformation from speech to song. Still, the determinants of the effect and the underlying mechanisms remain unclear. As shown for other massed repetition effects, phonetic properties of words, such as sound statistics, segmental acoustics, and articulatory representations, play a major role during perceptual transformations and illusions (Chandrasekaran & Kraus, 2010; Goldstein & Lackner, 1972; Kondo & Kashino, 2007; Pilotti, Antrobus, & Duff, 1997; Pitt & Shoaf, 2002; Sato, Schwartz, Abry, Cathiard, & Loevenbruck, 2006; Sato, Vallée, Schwartz, & Rousset, 2007; Warren & Meyers, 1987). For the S2ST, observed with longer utterances, prosodic features like intonation and rhythm are most promising features as determinants of the perceptual transformation. All prominent models that relate music and speech processing have identified prosody as the key element giving rise to the perception of musical or linguistic signals (Besson & Schön, 2001; Koelsch, 2012; Patel, 2003, 2008, 2011, 2012; Patel & Peretz, 1997; Peretz, 2012; Peretz & Coltheart, 2003; Zatorre & Gandour, 2008).

The prosodic systems of speech and song rely on the same acoustic resources, such as fundamental frequency (i.e., needed for pitch perception) and temporal patterns (i.e., needed for rhythm perception). They also widely share similar spectral properties resulting in syllable and word structure. Early perceptual processes like the analysis and encoding of periodicity may therefore be shared (Patel, 2008, 2011). Still, the higher-order structure and the functional value of acoustic resources seem to be processed in different ways (Dalla Bella, Berkowska, & Sowiński, 2011; Dalla Bella, Białuńska, & Sowiński, 2013; Peretz & Zatorre, 2005; Stewart, von Kriegstein, Dalla Bella, Warren, & Griffiths, 2009). As far as pitch is concerned, songs rely on a system of discrete pitch values that allows for the perception and production of interval structure, musical scales, as well as consonance and dissonance (Dalla Bella, Giguère, & Peretz, 2007; Dalla Bella & Berkowska, 2009; Krumhansl, 1990, 2000; Tan, Pfordresher, & Harre, 2010). In speech, on the other hand, exact pitch values are far less relevant for successful communication. Linguistically meaningful use of pitch, such as intonation or lexical tone, is based on relative pitch differences and is evaluated in the context of the actual utterance and in relation to the speaker's voice range (Fox & Qi, 1990; t'Hart, Collier, & Cohen, 1990; Nolan, 2003; Xu, 1994). Regarding rhythm, songs typically exhibit highly repetitive patterns of accents recurring with high temporal precision and involving layers of embedded periodicities (known as meter; see London, 2004; Tan et al., 2010). The resulting isochronous beat or pulse serves as a benchmark for accentual timing (e.g., Cooper & Meyer, 1960; Drake & Palmer, 1993). These aspects are typically not found in speech. Here, the accentual structure is temporally much more variable as a function of lexical and pragmatic meaning (see Arvaniti, 2009; Auer, Couper-Kuhlen, & Müller, 1999; Dalla Bella et al., 2013; Dauer, 1983; Nolan & Asu, 2009; Roach, 1982). In sum, processing of speech and song relies on the same acoustic resources, but processing demands may vary as a result of a different use of both rhythm and pitch structures to convey different communicative functions.

The S2ST provides a unique opportunity to examine the perceptual linkage between music and language using the same stimulus material. In this study, we seek to determine the prosodic properties of speech having the potential to activate musical processes and memory representations during repetition. In previous studies, it was shown that the perception and reproduction of pitch is prone to a substantial change during the S2ST (Deutsch et al., 2011). Therefore, it is likely that repetition enhances those features in the speech signal that can be interpreted as forming a musical melody. Still, we do not know which aspects of pitch are most relevant to foster the S2ST, an issue that will be addressed in the present paper. Furthermore, we investigate the role of rhythmic patterns in eliciting the transformation, a possibility that has not been addressed so far. Repetition is likely to induce temporal periodicities, thereby helping the listener to predict the timing of subsequent accent patterns (Cummins & Port, 1998; Cummins, 2009; Simko & Cummins, 2010). The present study provides the first systematic examination of the prosodic (i.e., pitch and rhythmic) characteristics that are most or least likely to induce the transformation. The effect of manipulating pitch and rhythmic properties on the frequency of the S2ST and on its moment of occurrence is tested.

We hypothesize that the transformation is achieved by a mechanism of functional reevaluation of prosodic properties. Those aspects relevant to speech processing will dominate perception at the beginning of stimulus presentation whereas the salience of the aspects relevant to song processing will increase during repetition. In two Experiments, we evaluated the role of pitch and rhythmic aspects of German sentences to account (a) for the emergence of the S2ST, and (b) for the number of repetitions needed to elicit the effect.

In naturally read spoken sentences, we varied the occurrence of prosodic features relevant for song or speech perception. We expect spoken sentences which contain cues to a musical interpretation to induce the transformation into song more often and after fewer repetitions than those which do not manifest them. Note that these musical cues (see below) may also occur in natural speech. However, they occur less frequently, less systematically and are not as functionally relevant in speech as in song processing.

The pitch and rhythmic aspects that were manipulated were derived from analyses on samples of naturalistic read speech eliciting the S2ST in previous studies (Deutsch et al., 2011; Tierney et al., 2012). As potential pitch cues to the S2ST, tonal target stability and scalar interval structure were chosen. Tonal targets were defined as local maxima and minima of fundamental frequency (f0). In speech, those targets are most often realized as dynamic patterns with little steady-state portions in their trajectories (e.g., Barnes, Veilleux, Brugos, & Shattuck-Hufnagel, 2010; Gussenhoven & Rietveld, 1998; t'Hart et al., 1990; Niebuhr, 2007; Ward & Hirschberg, 1985). In contrast, tonal targets in music are typically stable and define pitch classes and intervals necessary for melody recognition (e.g., Koelsch & Siebel, 2005; Tan et al., 2010; Zatorre, Belin, & Penhune, 2002). Figure 1 depicts the pitch contour of the original naturally read stimulus leading to the S2ST (Deutsch et al., 2011).

The phrase shows several relatively steady-state portions of its f0 trajectory, especially toward the end of the phrase. Additionally, it exhibits a descending eight-semitone interval (i.e., a minor sixth in music) on the final word. Deutsch et al. (2011) report that, after the perception of the S2ST, participants replicated the sentence's pitch contour with an elaborated interval structure. Therefore, the presence of a prominent interval may be important to elicit the S2ST by inducing a search for scalar interval structure in the whole speech signal. Note that in music, intervals represent fixed relations between discrete pitches that form a musical scale (e.g., seven pitches in the Western tonal system; Krumhansl, 1990, 2000, 2005). In speech, however, the scaling of tonal targets and their intervals vary considerably, for example, depending on discourse functions, degrees of prominence and other language-specific factors (e.g., Gussenhoven & Rietveld, 2000; Ladd & Morton, 1997; Lieberman & Pierrehumbert, 1984; Niebuhr, 2007; Rathcke, 2013).

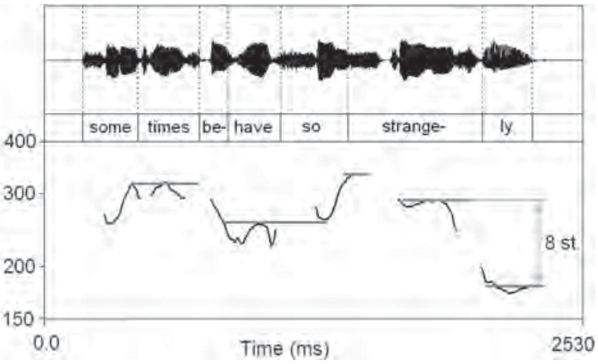


Figure 1. Waveform, syllabic labeling, and f0 trajectory of the English sentence (*but they) sometimes behave so strangely* used in the original experiment by Deutsch et al. (2011).

As potential rhythmic cues to the S2ST, regular accent distribution and intervocalic segmental grouping are particularly relevant. Rhythm perception in both speech and music has been described as resulting from a grouping mechanism that creates patterns of prominence recurring in time (e.g., Arvaniti, 2009; Clarke, 1999; Krumhansl, 2000). This mechanism requires the computation of accentual as well as temporal relations between basic events, such as notes, segments, syllables, or phrases. In previous studies, stimuli prone to the S2ST tended to have regular spacing of interaccent intervals (Tierney et al., 2012). A regular distribution of accents could increase a sense of rhythm in the speech signal and serve as a cue to musical beat perception (London, 2004). The present study examined whether a regular distribution of accented and unaccented syllables within a sentence would facilitate the S2ST. The timing at the level of the basic units of rhythmic structure may also affect the transformation. The basic rhythmic unit in speech and song seems to rely on different segmental grouping. Whereas intervocalic intervals are assumed to constitute the smallest units that convey rhythmic structure of songs (Sundberg, 1989), whole syllables are widely agreed to reflect the smallest rhythmic units of speech (Cutler, 1991). In casual spontaneous speech, syllabic intervals vary in duration depending on their prominence in the utterance (e.g., accented syllables are typically longer than unaccented ones; see Beckman & Edwards, 1990). In contrast, rhythmic units in songs may have the same duration whether they are accented or not. Accordingly, if intervocalic intervals are very regularly timed, this could facilitate the transformation into song during the course of repetition. These rhythmic properties were tested together with pitch properties in Experiment 1. In Experiment 2, the contribution of rhythmic properties was examined in more detail.

In addition to these prosodic properties, we also investigated individual predispositions in perceiving the transformation. The participants of the previous experiments on the S2ST were musicians with an average of 10 or more years of musical training (Deutsch et al., 2011; Tierney et al., 2012). So far, it is unknown whether the S2ST is a more widespread phenomenon, extending to listeners with less or no musical training. Musical experience may bias the listener toward perceiving musical structure in speech, and thus the S2ST, more readily. This question was also addressed in Experiment 1. Finally, the S2ST has until now only been demonstrated with English material. By showing that the effect arises with another language (i.e., German) differing in several prosodic aspects (e.g., inventory and realizations of pitch categories, Grabe, 1998; Mennen, Schaeffler, & Docherty, 2012; syllabic organization, Adsett & Marchand, 2010) we aim at testing whether the S2ST is a general perceptual phenomenon.

Experiment 1

Method

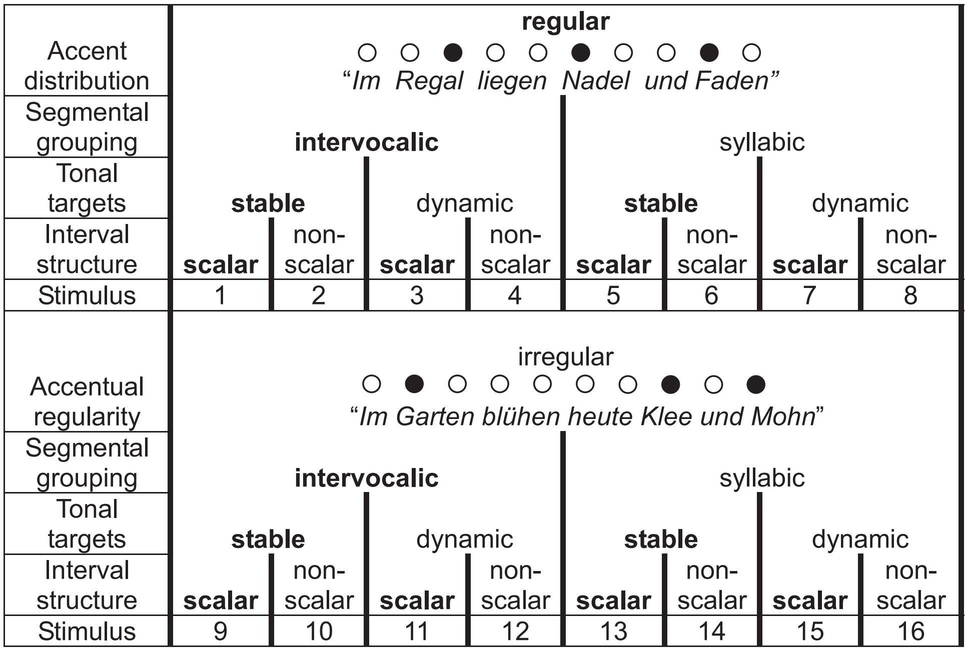
In Experiment 1, perceptual effects of pitch and rhythmic manipulations on the S2ST were examined. In particular, two pitch properties (tonal targets, interval structure) and two rhythmic properties (accent distribution, segmental grouping) were systematically varied in German sentences. Sentences containing prosodic cues relevant to musical processing were expected to facilitate the S2ST and to lead to faster emergence of the transformation than sentences that contained prosodic properties typical of casual spontaneous speech.

Participants. Sixty-two native German speakers (49 females), aged between 19 and 46 years (M = 24.0 years, SD = 4.9 years), volunteered to participate in Experiment 1. They were all undergraduate students at the Ludwig-Maximilians-University in Munich, and had on average 5.9 years (SD = 3.8 years) of musical training. Thirteen of the participants were nonmusicians (i.e., they had no or little musical training, less than one year).

Materials. To create the experimental material, four prosodic properties of sentences were manipulated, and two conditions for each property were obtained: Accent distribution (regular vs. irregular), segmental grouping (intervocalic vs. syllabic), tonal targets (stable vs. dynamic), and interval structure (scalar vs. nonscalar). The first condition of each property is assumed to be conducive to the S2ST and will be referred to as a "cue" to musical perception. Sentences with cues such as regular accent distribution, intervocalic segmental grouping, stable tonal targets, and scalar interval structures, are expected to facilitate and speed up the S2ST. In contrast, in the second condition, as these cues are absent (henceforth referred to as "no-cue"), the S2ST is less expected.

Two base sentences were chosen to implement regular (cue) versus irregular (no-cue) accent distributions. These sentences served as the basis for all further manipulations as illustrated in Figure 2. Overall, 16 test sentences were derived from the two base sentences, by progressively applying the manipulation of the other prosodic properties (i.e., segmental grouping, tonal targets, and interval structure) in a 2 × 2 × 2 × 2 design. For each prosodic property, eight sentences contained the acoustic cue related to potential song perception ("cue sentences"), and the other eight did not ("no-cue sentences").

The two base sentences consisted of 10 syllables each and had three lexically stressed syllables to be realized as accented in a broad focus reading. In the base sentence with regular accent distribution *Im Re**gal** liegen **Na**del und **Fa**den* ('There are needle and thread on the shelf'), every third syllable carried an accent (highlighted in bold). In contrast, the irregular base sentence, *Im **Gar**ten blühen heute* ***Klee*** und ***Mohn*** ('Today, clover and poppy are blooming in the garden') showed no regularity of accented and unaccented syllables. The segmental material of the selected base sentences was chosen to include mostly highly sonorant segments, thus allowing for detectable manipulations of the f0 contour. The base sentences were read by a female speaker of Standard High German and recorded in a sound-isolated booth at the Institute of Phonetics and Speech processing of the Ludwig-Maximilians-University in Munich. Subsequently, manipulations outlined in Figure 2 were applied using *Praat* Software (Boersma, 2001). Acoustic features of the manipulations are illustrated in Figure 3.



Accent distribution	**regular** ○ ○ ● ○ ○ ● ○ ○ ● ○ *"Im Regal liegen Nadel und Faden"*							
Segmental grouping	**intervocalic**				syllabic			
Tonal targets	**stable**		dynamic		**stable**		dynamic	
Interval structure	**scalar**	non-scalar	**scalar**	non-scalar	**scalar**	non-scalar	**scalar**	non-scalar
Stimulus	1	2	3	4	5	6	7	8
Accentual regularity	irregular ○ ● ○ ○ ○ ○ ○ ● ○ ● "*Im Garten blühen heute Klee und Mohn*"							
Segmental grouping	**intervocalic**				syllabic			
Tonal targets	**stable**		dynamic		**stable**		dynamic	
Interval structure	**scalar**	non-scalar	**scalar**	non-scalar	**scalar**	non-scalar	**scalar**	non-scalar
Stimulus	9	10	11	12	13	14	15	16

Figure 2. Schema of material preparation. The manipulated properties are indicated on the left, base sentences and conditions on the right. Cue sentences are indicated in bold.

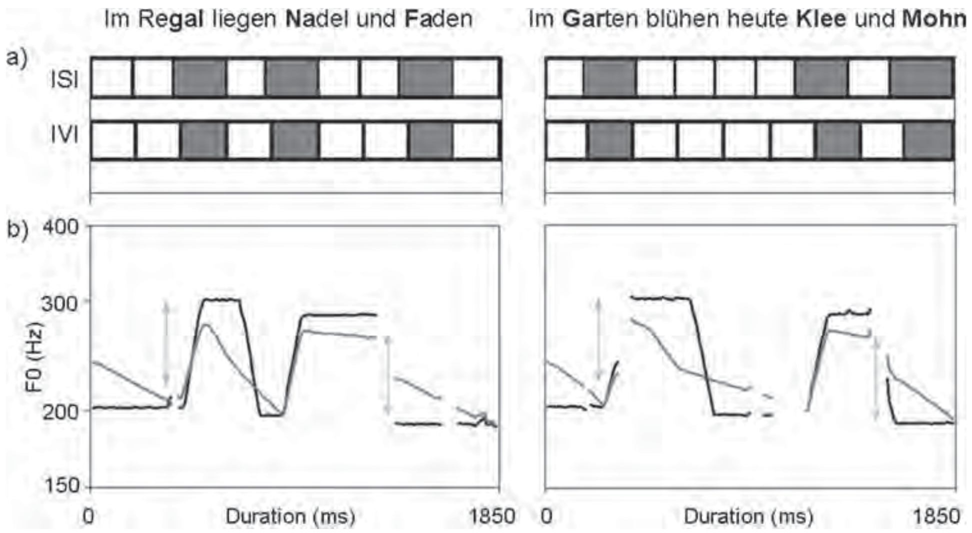


Figure 3. Pitch and rhythm manipulations for accentually regular base sentences (left), and irregular sentences (right). (a) Illustration of durational intervals in sentences with intersyllabic grouping (ISI) and intervocalic grouping (IVI). Accented intervals are indicated in bold, their duration (in ms) is represented by gray boxes. (b) Manipulations of tonal targets in sentences with stable tonal targets and scalar interval structure (black lines) and sentences with dynamically changing pitch contours and without musically significant intervals (gray lines). The arrows refer to the position of the critical intervals on the first and last accented syllable.

To test the role of segmental grouping, the durational structure of the base sentences was varied at the level of intervocalic or intersyllabic intervals (IVI vs. ISI; Figure 3a). Nine[1] IVIs/ISIs were computed for each base sentence. IVIs were obtained by measuring the duration between two nucleus onsets. Syllabic boundaries, needed to compute ISIs, were determined following the maximal onset principle (Selkirk, 1981). In test sentences with intervocalic grouping structure (cue sentences), all nonfinal IVIs were set to be equal in duration (200 ms). In test sentences with syllabic grouping, the duration of the nonfinal ISIs varied depending on their prominence: unaccented syllables were shorter (150 ms) than accented ones (250 ms). To ensure that the sentences sounded natural, phrase-final lengthening was implemented on the last and/or penultimate interval depending on accent structure (Lehiste, 1973; Lindblom, 1978; Turk & Shattuck-Hufnagel, 2007). The total duration of each test sentence was 1850 ms.

Tonal targets were varied starting from the speaker's original pitch contour and range produced in both base sentences (see Figure 3b). A declination of 0.5 semitones was applied to both top- and baseline (to reflect the natural tendency of pitch to drift downward during an utterance, e.g., t'Hart & Cohen, 1973). In test sentences with stable tonal targets (cue sentences), f0 contours between tonal targets were flattened whereas in test sentences with dynamic targets (no-cue sentences), the contour changed dynamically with respect to segmental landmarks.

Finally, interval structure was manipulated by introducing either a scalar interval (in cue sentences) or a nonscalar interval (in no-cue sentences) in the sentences. The interval appeared twice in each sentence: on the first accented syllable (ascending interval), and on the last accented syllable (descending interval). The scalar interval was a 7-semitone interval (i.e., perfect fifth), one of the most prominent and consonant intervals in musical scales around the world (Krumhansl, 1990; Schellenberg & Trehub, 1996). The pitch distance of the nonscalar interval was set to 5.5 semitones, which does not correspond to the musical scale of the Western tonal system.

To avoid continuous presentation of the same semantic and segmental content across the test session, filler sentences were introduced. The fillers were based on four sentences with a syntactic and rhythmic structure similar to the base sentences. They differed from the test sentences semantically, morphophonologically and partly with respect to the number of accents. Sixteen fillers were created (4 sentences × 4 prosodic manipulations). Manipulations concerned the duration of IVIs/ISIs, the duration of the sentence, as well as tonal targets. Overall, eight fillers had stable and eight fillers had dynamic tonal targets.

Procedure. The overall 32 sentences used in Experiment 1 (16 test sentences, 16 fillers) were divided into two sets, each containing eight test sentences (four cue and four no-cue sentences) and eight fillers. A subgroup of participants ($n = 30$) selected randomly from the overall group were tested with one set; the other group ($n = 32$), with the second set. In one trial, each sentence was looped with 10 repetition cycles and a 400-ms pause between each cycle. Note that the pause was shorter than in previous studies (700 ms in Deutsch et al., 2011, and 500 ms in Tierney et al., 2012). Our choice of a shorter pause was motivated

[1] Although we had 10 syllables per sentence, only nine durational intervals were calculated. This was attributable to German vowel reduction that emerged during material preparation. To avoid that the sentences sounded unnatural, we treated phrase-medial reduced forms as single rhythmic units, that is, [liːgn], [blyːn], but phrase-final [faːdn̩] as two different units. This decision was motivated by previous findings that rhythmic units in phrase-final positions tend to have additional duration as compared to phrase-medial positions (e.g., Lehiste, 1973).

by a pilot experiment, showing that a shorter duration between cycles resulted more often in the S2ST than when the pause was longer.

Participants were tested individually in a quiet room. We adopted the procedure and instructions used in previous studies with a slight modification that enabled us to collect data of the moment of occurrence of the S2ST (Deutsch et al., 2011; Tierney et al., 2012). As done in studies of other well-established perceptual transformations, such as auditory stream segregation (e.g., Bregman, 1990; Helenius, Uutela, & Hari, 1999), participants were instructed about alternative perceptual impressions they might experience during the experiment (i.e., that after multiple repetitions, a sentence might sound either spoken or sung). In a pilot study, we found that listeners tended to report a variety of perceptual transformations (including the S2ST) when no explicit instruction was provided.[2]

An experimental session started with three practice trials. The original English phrase (taken from Deutsch, 2003) was presented along with two German sentences that proved effective in inducing the S2ST in the pilot study. Participants listened to the looped sentences in each trial. As soon as they had the impression that the sentence was no longer spoken, but sounded like song, they pressed a key (Enter) of the computer keyboard. The response stopped the trial. Following the response, or after 10 repetition cycles if there was no response, participants confirmed by another key press whether or not they had perceived a change. After each trial, participants were asked to solve a simple mathematical equation. This task served to distract the participants' attention from the presented sentence, to clear working memory, and to avoid carry-over effects of the memory traces to the following trial. To determine whether individual differences in music perception skills may affect the S2ST, participants were additionally tested on their ability to discriminate melodies (with the scale subtest of the Montreal Battery of Evaluation of Amusia; Peretz, Champod, & Hyde, 2003). In this task, participants were asked to compare two melodies presented in succession ($n = 30$), and to decide whether the melodies were the same or different; in half of the cases, the second melody differed from the first by one note.

Experiment 1 was run on an IBM-compatible computer using DMDX software (Forster & Forster, 2003). Written instructions were given on the computer screen at the beginning of each experimental session. Sentences were presented in pseudorandomized order respecting the following rules: (a) a session always started with a filler and (b) at least two semantically different (test or filler) sentences intervened between two test sentences with the same semantic content. The sound material was played back via Sennheiser HD-424 headphones at a volume comfortable for the participant. Experiment 1 lasted between 15 and 20 minutes.

Results and Discussion

Before data analysis, trials with false starts (i.e., when participants responded during the first presentation of the sentence) were discarded (i.e., 14 trials, 5% of the dataset). Only the results obtained with test sentences were submitted to further analyses. Below, we will report the results for (a) the probability of occurrence of the S2ST and (b) the repetition cycle at which the transformation was reported.

Fifty-nine of 62 participants reported the S2ST. On average, these participants perceived the transformation in 65% of the test sentences with individual differences ranging from 25% to 100% ($SD = 27.2\%$). Mean perception of the S2ST per test sentence averaged across participants ranged from 37% up to 83% ($M = 65\%$, $SD = 14.7\%$). Figure 4 shows the probability of perceived transformations for each property averaged across participants (expressed in percentage).

To test the predictive power of the manipulated properties on the probability of occurrence of the S2ST, the results from all cue sentences for a given property were pooled together and compared with the no-cue sentences for the same property, irrespective of the remaining properties. The data were entered into a generalized linear model (binomial) to perform an ordinary logistic regression using R software environment (version 2.13.0). The dependent variable was dichotomous, with "1" representing the positive S2ST-response and "0" the negative S2ST-response per participant/sentence. Each predictor (i.e., Accent distribution, Segmental grouping, Tonal targets, and Interval structure) was treated as a dichotomous variable, with value 1 for the cue sentences and value 0 for the no-cue sentences. The best fit of the model was tested using the Likelihood Ratio Test ($\chi^2 = 37.66$, $p < .001$).

The results revealed that pitch, but not rhythmic properties, affected the perception of the S2ST. Target stability was the most powerful cue to facilitate the S2ST ($p < .001$). Test sentences with stable tonal targets evoked the transformation 25% more often than sentences with dynamic targets. Furthermore, there was a marginally significant effect for test sentences with scalar intervals which were slightly more often reported to transform into songs in comparison to the sentences with a nonscalar interval structure ($p = .096$). No further main effects or interactions were found.

With respect to speed of occurrence, the S2ST emerged most frequently during the third repetition cycle (of 10). The data for repetition cycle (cycle 2 to 10, only positive answers = 65% of the test sentences) were similarly entered into a generalized linear model (Gaussian) taking the same four predictors described above (best fitting model, Likelihood Ratio Test, $\chi^2 = 6.93$, $p < .001$). For pitch properties, the results replicate the results of the analyses presented above. The S2ST occurred earlier if sentences had f0 contours with stable tonal targets (mean cycle = 4.4, $SD = 1.7$) in contrast to dynamic targets (mean cycle = 5.3, $SD = 2.0$). The effect was highly significant ($p < .001$). The difference between scalar and nonscalar interval structure just failed to reach significance ($p = .08$). For rhythmic properties, however, a main effect was found for segmental grouping ($p < .01$). The best fit model further contained an interaction of segmental grouping and accent distribution ($p < .01$). Sentences with syllabic intervals induced the transformation earlier when they were embedded in sentences with regular accent distribution ($M = 4.6$, $SD = 2.1$) in contrast to sentences with irregular accent distribution ($M = 5.3$, $SD = 2.1$).

[2] In the pilot experiment, we asked participants to report whether the repeated stimulus was changing over time and what was changing. With this instruction, the S2ST was indeed perceived by some participants, but other participants reported different kinds of percepts such as verbal transformations, illusory tempo, and amplitude changes. Therefore, in the main Experiments, we opted, as done in previous studies, for more targeted instructions to have comparable results among participants.

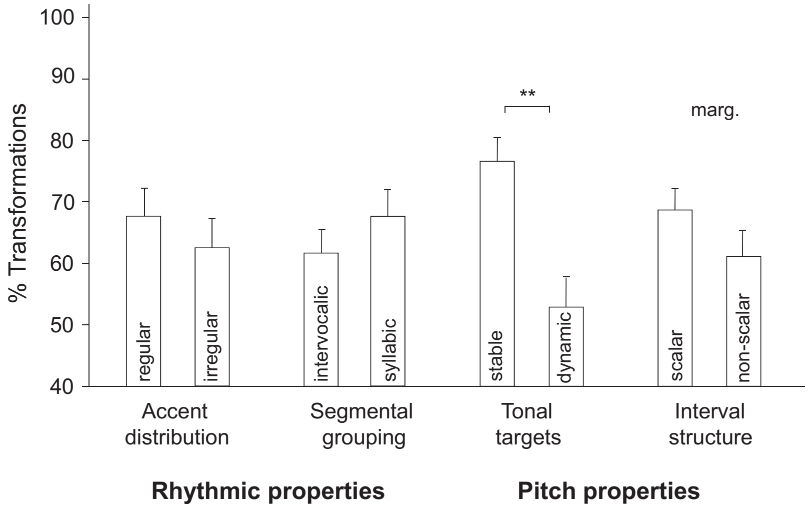


Figure 4. Percentage of perceived S2ST as a function of the four properties manipulated in Experiment 1. ** $p < .001$; marg. = marginally significant. Error bars display the standard error of the mean (averaged over participants).

As for individual differences, participants who reported more occurrences of the S2ST also tended to perceive the transformation after fewer cycles within the trial ($r = -.38$, $p < .01$). This result cannot be merely attributed to an effect of practice with the task, because participants did not progressively speed up their response during the course of the Experiment. Additionally, participants with more years of musical training tended to perceive the S2ST later in the loop as compared to less musically trained individuals ($r = .26$, $p < .01$). No further differences were found between musicians and nonmusicians. Both groups were equally able to experience the S2ST. Musical training was also positively correlated with the performance in the Montreal Battery of Evaluation of Amusia ($r = .34$, $p < .01$). However, no relation was found between the results in the MBEA and individual performances in the S2ST.

As predicted, prosodic properties of spoken phrases affected the occurrence and the speed of the S2ST. Our findings support the idea that pitch aspects play a major role in facilitating the transformation (Deutsch et al., 2011). Participants perceived the transformation more frequently and at an earlier repetition cycle when the pitch contour was constituted by stable tonal targets. This result is consistent with the comparative literature on speech and music showing that pitch is the most salient cue to musical structure (Besson & Schön, 2001; Patel, 2008; Peretz & Coltheart, 2003). However, some cues (i.e., stable tonal targets) were much more efficient in eliciting the S2ST than others (i.e., prominent musical intervals).

Rhythmic cues had little effect on the S2ST. They only affected the time course of the transformation, not its probability of occurrence. The time course was influenced by a combined effect of accent distribution and segmental grouping. A regular accent distribution facilitated the perception of the S2ST in sentences with syllabic grouping compared with an irregular accent distribution. In addition, pitch and rhythmic cues did not interact.

In contrast to our initial hypothesis, a facilitating effect of accentual regularity was not visible in sentences with intervocalic grouping structure. Still, the syllabic grouping condition was supportive of a musical interpretation. The syllabic grouping condition featured durational contrasts of longer accented and shorter unaccented syllables. This rhythmic property may have made the longer, accented syllables more salient and thereby facilitated the perception of the regular prominence pattern. In a similar vein, the equalized duration of intervocalic intervals may have hindered the perception of a clear prominence pattern.

Two further aspects deserve to be examined to further assess the effect of rhythmic properties on the S2ST. Because most Western tonal music is based on an isochronous pulse, a high temporal regularity of accents in speech might strengthen the impression of musical beat structure, thus facilitating the S2ST. In our material, accentual regularity was implemented as structural regularity (i.e., an accent every three syllables) but not as temporal regularity (i.e., the durational interval between accents was still varying). This derives from the fact that phrase medial schwa vowels were reduced to a single nasal nucleus (i.e., [liːgn], [blyːn]) and had to be chunked into one temporal unit together with the preceding syllable to sound natural. Therefore, in test sentences with regular accent distribution, the first and second accent were temporally slightly closer than the second and third accent. Finally, the regularity of timing of prominence patterns across repetition cycles may also be a relevant rhythmic aspect. For example, looping of auditory events is widely exploited in contemporary electronic music in order to create musical beat structures (Souvignier, 2003). The loop-internal timing is related to the duration of the pause between repetition cycles. Pause durations allowing for a high temporal regularity of accent occurrence across the whole loop (i.e., approaching an isochronous pattern of accents) may also facilitate the S2ST. In contrast, longer or shorter pauses should be comparatively less effective.

The next Experiment was conducted to further examine the potential contribution of rhythmic properties to the emergence of the S2ST, by implementing the aforementioned modifications of the test materials. We hypothesized that a prominence pattern that is (a) conveyed by clear accentual differences and (b) temporally regular within and across repetitions should facilitate the S2ST. An additional goal of Experiment 2 was to test whether individual differences in terms of temporal processing, as revealed by a sensorimotor synchronization task, could account for the individual speed and probability of occurrence of the S2ST. Again, this would be indicative of an important role of rhythm in the transformation.

Experiment 2

Method

In Experiment 2, three rhythmic properties (accent distribution, segmental grouping, and pause duration between repetitions) were manipulated to clarify the role of rhythmic structure in the S2ST. Rhythmic cues that create higher levels of accentual regularity and highlight the underlying beat structure, a typical trait of musical signals, were expected to enhance the likelihood of occurrence of the S2ST. We implemented two properties previously examined in Experiment 1 (accent distribution and segmental grouping) with slight modifications to the experimental material. Additionally, Experiment 2 tested one previously unexamined property, the role of the pause between repetitions.

Participants. Thirty native speakers of German (23 females), all undergraduate students at the Ludwig-Maximilians-University in Munich, aged between 19 and 32 years (M = 24.3 years, SD = 3.3 years), volunteered to participate in Experiment 2.

Materials. Two new base sentences were selected. They had only unreduced vowels in both accented and unaccented syllables to prevent vowel elision in unaccented syllables. Again, the base sentences were formed by 10 syllables. The sentences started with an accented syllable, and had four accents in total (not three as in Experiment 1) to better implement a regular accent distribution within and across sentences (see below). They were recorded by another female speaker at the Institute of Phonetic and Speech processing of the Ludwig-Maximilians-University. The original f0 trajectories were stylized so as to create comparable pitch contours in the two base sentences and to control for pitch height. The resulting contour shapes were kept constant across all subsequent manipulations. Manipulations were done in a similar way as described in Experiment 1 (see Figure 2). Accent distribution and segmental grouping were manipulated by obtaining two conditions for each property, one including a cue theoretically conducive to the S2ST (cue sentences), and the other without such a cue (no-cue sentences). Pause duration had three conditions, one including the cue (matched pause) and two without the cue (shortened, lengthened pause). This procedure resulted in 12 test sentences (2 × 2 × 3 manipulations).

The accentually regular base sentence was ***Ti****na will* ***nie****mals im* ***Ju****li nach* ***Rom*** ("Tina would never go to Rome in July"). Each accented syllable was followed by two unaccented ones. In test sentences derived from this base sentence, the duration between accents[3] was set to 600 ms in the beginning and then slightly increased toward the end of the sentence (SD of the interaccent intervals = 58 ms). This was done to account for the fact that, unlike in music, temporal regularity in linguistic stimuli is best perceived when intervals between accents become longer throughout a sentence (Lehiste, 1973; Nakajima, ten Hoopen, & van der Wilk, 1991; Sasaki, Suetomi, Nakajima, & ten Hoopen, 2002). The accentually irregular sentence was ***Tim*** *will im* ***Mai ger****ne in* ***Berlin sein*** ("Tim would like to be in Berlin in May"). The accented syllables appeared irregularly and there was a stress clash between the second and third accent (***Mai ger***ne). This structure was expected to disrupt any perception of temporal regularity. In test sentences derived from this base sentence, the duration between accents was variable (SD of the interaccent intervals = 231 ms; see Figure 5).

Segmental grouping was manipulated by creating chunks in terms of ISIs and IVIs as described in Experiment 1. However, in contrast to Experiment 1, durational manipulations were done in such a way that both accented ISIs and IVIs were 24% to 40% longer than unaccented intervals and increased in duration toward the end of the sentence to reflect phrase-final lengthening (e.g., Turk & Shattuck-Hufnagel, 2007, see Figure 5). Duration ratios from the natural production of the base sentences served as the basis of the manipulation. Furthermore, ISIs and IVIs were matched with each other in duration. These manipulations led to a total duration of each test sentence of 2400 ms.

Finally, three different durations of the pause intervening between repetitions were implemented (see Figure 5). A highly regular loop was obtained for accentually regular sentences by setting the pause to 350 ms. In this case, the duration of the accentual interval between repetition cycles matched the duration of accentual intervals within the sentence (matched condition). Shortening and lengthening the pause by 180 ms resulted in a faster and slower succession of cycles and was meant to undermine the impression of regularity. In accentually irregular sentences, pause durations were the same as in the accentually regular ones (350 ms ± 180 ms). However, no perception of regularity was expected across those repetitions because of the internal temporal irregularity of the base sentence.

As in Experiment 1, filler sentences were added. The fillers were based on the recordings of four additional sentences which were structurally similar to the test sentences. To make the prosodic variants of all experimental material comparable, three manipulations were applied to each filler sentence. The manipulations included stylization of pitch contour, and shortening or lengthening of the duration of accented or phrase-final syllables. Moreover, different pause durations were used. The procedure resulted in 12 fillers.

Procedure. The task and equipment in Experiment 2 were the same as in Experiment 1, with two changes. The overall set of 24 sentences (12 test sentences, 12 fillers) was looped 7 times (instead of 10 as done in Experiment 1). Fewer repetitions were judged to be sufficient, because in Experiment 1 the transformation was perceived mostly as soon as in the third repetition. Each participant was tested on all 24 sentences presented in three blocks of eight sentences each (four test sentences, four fillers per block). The first block started with two fillers; block 2 and 3 always started with one filler sentence at least. The remaining sentences within a block

[3] The measurement of the accented interval was dependent on segmental grouping. This interval was measured either from the beginning of the syllable (syllabic grouping) or from the beginning of the accented vowel (intervocalic grouping).

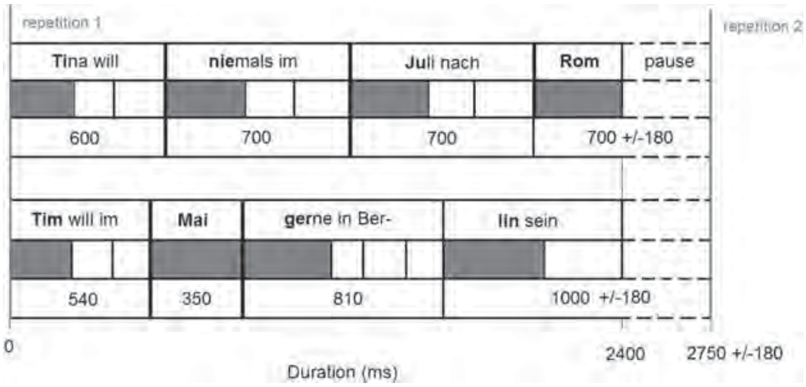


Figure 5. Temporal manipulations of the accentually regular (top) and irregular (bottom) base sentences within and across repetitions in Experiment 2. Accented intervals are marked in bold, their duration (in ms) is represented by gray boxes. Only syllabic grouping is shown. The durational and accentual structure of intervocalic intervals was exactly the same.

were presented in random order. In addition, participants' abilities in temporal processing were assessed using a sensorimotor synchronization task (Aschersleben, 2002; Repp, 2005, 2006). Participants tapped with their index finger to 100 tones presented isochronously (Inter-Onset-Interval, IOI = 600 ms). Tapping was recorded with a Roland SPD-6 electronic percussion pad. The experimental material was presented and data collected with Sonar 3 software. The finger tapping task was performed three times, before the test session on the S2ST, and after the first and the second block. Experiment 2 lasted approximately 15 minutes.

Results and Discussion

Data from 27 trials including false starts (see Experiment 1, 8% of the overall data set) were discarded. Two participants showing more than 50% false starts were also excluded from further analysis. All remaining 28 participants reported the S2ST in the test sentences. On average, participants experienced the S2ST with 76% of the test sentences (range = 30–100%; *SD* = 19.8%). Mean perception of the S2ST per test sentence averaged across participants ranged from 41% to 100% (*M* = 78%, *SD* = 15.9%). The percentage of perceived transformations as a function of the three rhythmic properties under investigation is displayed in Figure 6.

As done in Experiment 1, the yes/no-responses of participants on their perception of the S2ST were entered into a generalized linear model (binomial) to perform an ordinary logistic regression (log-likelihood of the final model was $\chi^2 = 5.08$, $p < .05$). The predictors were Accent distribution, Segmental grouping, and Pause duration. The model revealed significant effects for two main predictors: Segmental grouping ($p < .001$) and Pause duration (shortened vs. matched: $p < .05$). Moreover, there was a significant interaction between Segmental grouping and Accent distribution ($p < .05$), indicating that segmental grouping had a stronger effect if there was a regular distribution of accents within the sentence. The percentage of S2STs was different for accentually regular and irregular test sentences only when intervocalic segmental grouping was present ($p < .05$). Interestingly, the matched pause inducing a regular accent occurrence across repetitions did not facilitate the transformation. Still, shortening the pause induced significantly more transformations than lengthening or matching the pause duration. As in Experiment 1, listener responses were also analyzed in terms of the speed of occurrence of the S2ST. The data from repetition cycle (only positive answers = 76% of the test sentences) were entered into a generalized linear model analysis (Gaussian model) with the same predictors as above. The model was nonsignificant (Likelihood Ratio test, $\chi^2 = 8.59$, $p > .05$). As observed in Experiment 1, irrespective of the manipulation of rhythmic properties, the transformation occurred most often at repetition cycle 3.

To assess the role played by individual differences in temporal processing, each participant's sensitivity for the rhythmic cues which may facilitate the S2ST was calculated. Sensitivity was defined as the difference between the individual percentage (and average cycle) of the S2ST for no-cue sentences and cue sentences, calculated for each rhythmic property. Each participant's sensitivity to the prosodic cues was correlated with accuracy in the synchronization task. Synchronization accuracy was measured by absolute mean asynchrony (i.e., the average absolute asynchrony of each tap with regard to the metronome beat pacing the participant). Small absolute asynchrony indicates high accuracy.

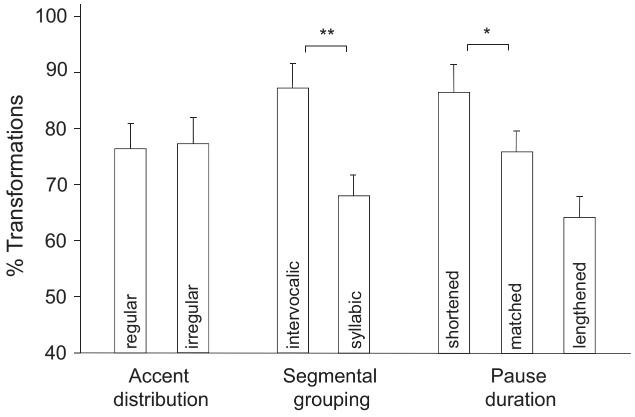


Figure 6. Percentage of perceived S2ST as a function of the three rhythmic properties manipulated in Experiment 2. $^{**}p < .001$; $^{*}p < .05$. Error bars display the standard error of the mean (averaged over participants).

Participants perceiving more transformations in accentually regular than irregular sentences were also more accurate in the synchronization task (i.e., they exhibited lower absolute mean asynchrony, $r = -.53$, $p < .01$). This finding indicates that a regular rhythmic structure of spoken utterances can serve as a cue to the transformation depending on the rhythmic abilities of a listener. In other words, individuals with fine-grained temporal processing capacities may be more affected by regular rhythmic structures in speech and use them as a perceptual cue to the S2ST.

To sum up, the results of Experiment 2 supported our hypothesis that rhythmic properties affect the occurrence of the S2ST. Segmental grouping was the strongest predictor of the S2ST. This is in line with the idea that intervocalic grouping conveying a clear pattern of prominence (accented/long vs. unaccented/short IVIs) is more effective in evoking the transformation than comparable syllabic grouping (Sundberg, 1989). Accent distribution had an effect on the S2ST, but only for sentences with intervocalic grouping (i.e., higher regularity within these sentences increased the probability of the transformation). Yet, contrary to our initial prediction, the manipulation of the pause between repetitions allowing for regular accent occurrence (i.e., matched to the rhythm within a sentence) was not the most efficient condition for evoking the S2ST. Instead, a fast succession of repetitions (i.e., a shortened pause) was found to induce more transformations.

Additional Tests

Two additional tests were conducted to better understand the nature of the results in Experiment 1 and Experiment 2. First, we wanted to ensure that the results found in this study were attributable to a genuine transformation from speech to song and not only the consequence of an initial evaluation of the sentence as sounding either spoken or sung. Therefore, base ratings for each test sentence were assessed and correlated with the results on transformation probability and speed (see Base ratings below). Second, as segmental grouping and regular accent distribution in Experiment 2 were not independent in affecting the occurrence of the S2ST, we assessed the nature of this interaction more thoroughly. We examined whether the perceived rhythmic pattern of accented and unaccented intervals in Experiment 2 differed in sentences with intervocalic versus syllabic grouping. The perception of rhythm in these sentences was tested in a new group of participants using a synchronized finger-tapping task (see Tapping test below). This task has been shown to be particularly informative about the perception of temporal properties in rhythmically simple and complex auditory stimuli, such as a metronome or music (for reviews, see Repp, 2005, 2006). It has also proven useful to examine rhythm perception with speech stimuli (Lidji, Palmer, Peretz, & Morningstar, 2011; Villing, Repp, Ward, & Timoney, 2011).

Base Ratings

To ensure that all test sentences from Experiment 1 and 2 were perceived as speech from the beginning, we assessed the way listeners interpreted them after a single presentation.[4] Eleven undergraduate students ($M = 26.6$, $SD = 5.2$, range = 23 to 42 years) who were unfamiliar with the S2ST were asked to rate the sentences (fillers of both Experiments included) as sounding like speech or song on a 7-point Likert scale (1 = *sounding clearly like speech*, 7 = *sounding clearly like song*). The sentences were presented in random order via headphones using DMDX software (Forster & Forster, 2003).

Results (see Figure 7) confirmed that all test sentences presented once were perceived as speech from the beginning ($M < 3$ for all conditions). A correlation between base ratings and the mean percentage of occurrence of S2ST per sentence did not reveal significant results. Therefore, we exclude that the probability of occurrence of the S2ST is linked to the initial perception of the sentences. On the other hand, the correlation between base ratings and mean repetition cycle was highly significant ($r[28] = -.713$, $p < .001$). This finding suggests that, if perceived, the transformation is likely to occur later with sentences that are initially judged as typical examples of spoken sentences.

In sum, these results confirm that the results obtained in Experiment 1 and Experiment 2 were attributable to a perceptual transformation from speech to song. Furthermore, the probability of the S2ST is independent of the initial perception of the sentence. Still, the speed of the S2ST occurrence is related to the perceived speech or song qualities of the sentence.

Tapping Test

Six right-handed undergraduate students from the Ludwig-Maximilians-University ($M = 21.8$ years, $SD = 3.2$, range = 19 to 28 years) performed a synchronized finger-tapping task. All of them were unfamiliar with the S2ST. Participants were asked to tap with the index finger of their dominant hand to every syllable of the accentually regular sentences used in Experiment 2, which were presented in both intervocalic and syllabic grouping conditions. Each sentence was repeated 14 times, with three different pause durations between repetitions (see Experiment 2). The testing was carried out with the same equipment used for the sensorimotor synchronization task in Experiment 2. The mean inter-tap-interval (ITI, in ms), averaged across participants and pause conditions for syllabic and intervocalic grouping, is presented in Figure 8a for stimuli with syllabic versus intervocalic grouping.

ITIs for the accented intervals were significantly longer than for the unaccented ones in sentences with intervocalic grouping (on average, 276 ms vs. 212 ms, respectively; Mann–Whitney $U = 0$, $p < .05$). Note that the duration of these ITIs closely matched the acoustic durations of IVIs (see Figure 8b) in the speech signal ($r = .82$, $p < .01$). In contrast, tapping results for sentences with syllabic grouping revealed a different rhythmic pattern. In this case, participants tapped in a quite regular fashion, not showing any difference in ITI between accented and unaccented intervals (with mean ITI = 237 ms in both cases, see Figure 8a). In addition, these ITIs were unrelated to acoustic durations of either ISIs or IVIs (Figure 8b).

The results obtained in the tapping task support our idea that the variation in segmental grouping led to systematically different interpretations of the rhythmic pattern in the sentence with otherwise identical linguistic content. Sentences with syllabic grouping evoked a rhythmic pattern that represents a sequence of similar durational units. On the other hand, sentences with intervocalic

[4] Obviously, the pause manipulation could not be included in the base ratings as these were collected by single presentations of sentences.

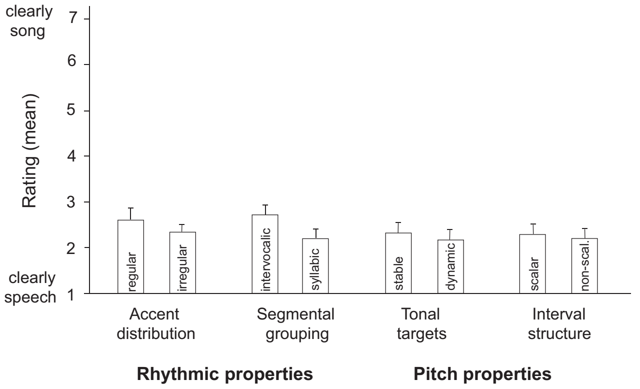


Figure 7. Base ratings of the test sentences without repetition per condition and property from Experiments 1 and 2. Error bars display the standard error of the mean.

grouping were perceived as including an important durational contrast between accented and unaccented intervals (e.g., with a long-short-short pattern). Moreover, the perceived contrast in these sentences is reminiscent of a 1:2 duration ratio which is frequently found in binary meters in music. Durational contrast is a key element (together with accentuation) in conveying the perception of a metrical pattern in music (e.g., London, 2004). This may be one of the main reasons why sentences with intervocalic grouping were more effective in evoking the S2ST. In general, metrical patterns in music involve nested periodicities of durational units (Lehrdahl & Jackendoff, 1983; London, 2004). The presence of subdivisions within a measure provided by lower periodicities also supports the perception of prominent events at higher levels. Therefore, in sentences with intervocalic grouping, the perceived durational subdivision of inter-accent-intervals reminiscent of a common musical ratio may have supported greater beat salience which was expected to be more conducive to the S2ST (see Discussion of Experiment 1).

Nevertheless, further reasons for why durational relations were perceived so differently as a result of segmental grouping remain unclear. Particularly in sentences with syllabic grouping, it is not excluded that some kind of perceptual compensation for contextual influences might take place, a phenomenon often attested in speech perception (e.g., Harrington, Kleber, & Reubold, 2008; Lehiste, 1973). This issue will need further enquiry in future studies.

General Discussion

In two Experiments, we investigated a massed repetition effect inducing a perceptual transformation from speech to song (the S2ST, Deutsch et al., 2011). Our first aim was to examine the role of acoustic structure of spoken phrases in eliciting the transformation. The second aim was to examine individual differences accounting for the perception of the transformation. The results clearly demonstrated that prosodic properties of spoken phrases were key determinants of the S2ST. Moreover, some dimensions of those properties were more important than others. Overall, pitch properties were found to be very reliable cues to the transformation. The presence of stable tonal targets in spoken utterances was the property most conducive to the S2ST (Experiment 1). In contrast, the occurrence of scalar intervals was not equally facilitating (Experiment 1). Notably, we document for the first time that rhythmic cues enhancing prominence contrasts facilitated the S2ST (Experiment 2). Sentences with intervocalic segmental grouping evoked the S2ST more often and earlier when temporal relations between accented and unaccented intervals approximated simple duration ratios and were regularly recurring (Experiment 2). However, a mere regularity of accent occurrences within and across sentence repetitions, a feature characteristic of music perception, was not in itself sufficient to facilitate the transformation (Experiments 1 & 2). Moreover, the proximity of repetitions in time enhanced the chance of perceiving the S2ST (Experiment 2). Finally, we showed that listeners' sensitivity to rhythmic cues could predict whether the S2ST was more or less likely to occur.

These results support the idea that the S2ST arises from a perceptual reevaluation of prosodic features during repetition. It is likely that listeners' attunement to fine-grained *relations* between pitch as well as rhythmic events in spoken utterances is enhanced during the course of repetition. In the present study, the S2ST was found to be cued most efficiently by prosodic features favoring a detailed representation of pitch trajectories (based on stable tonal

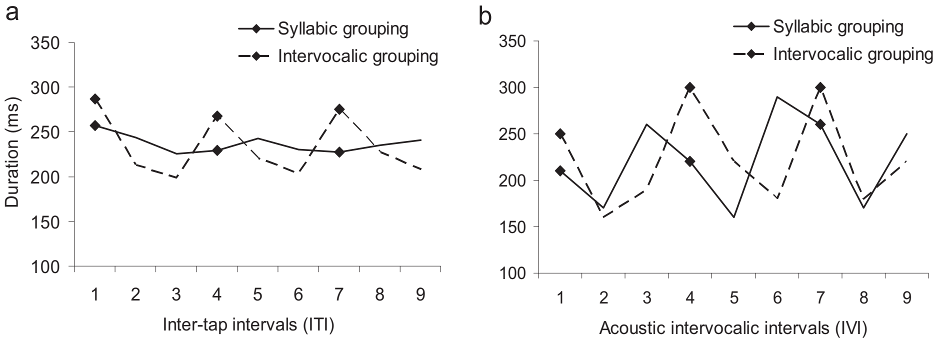


Figure 8. Tapping performance (a) and acoustic structure (b) of accentually regular test sentences (*Tina will niemals im Juli nach Rom*) with syllabic and intervocalic grouping structure. (a) Duration of nine intertap-intervals (ITI, in ms) as obtained in the tapping test. (b) Acoustic structure of the sentences measured in intervocalic intervals (IVI, in ms). Accented syllables are marked with a rhomb. The last interval was not taken into account because of the variable pause that was following.

targets) and prominence relationships (based on recurring durational contrasts between accented and unaccented events). It is worth noting that stable tonal targets are critical for the perception of discrete pitch and for the encoding of small differences between adjacent pitches (e.g., one or two semitones, Zatorre & Baum, 2012). Moreover, the temporal contrasts in our stimuli approximated small integer ratios forming rhythmic patterns which are typically best remembered and reproduced (Krumhansl, 2000). That musical stimuli are processed in more detail in these dimensions than verbal stimuli is shown when listeners are asked to discriminate fine-grained pitch relations and estimate metrical regularity (e.g., t'Hart, 1981; Lehiste, 1970; Tillmann et al., 2011). This is in keeping with the recent proposal that musical processing demands higher effort in encoding these features than speech processing (see the OPERA-framework, Patel, 2012). During the S2ST, higher activation of cortical areas associated with pitch and motor processing was also found and interpreted as reflecting higher processing demands (Tierney et al., 2012). In sum, massed repetition may foster the emergence of musical percepts by facilitating a more detailed representation of prosodic relationships, typically not required in a speech context. Nevertheless, in our view, detailed prosodic encoding cannot alone account for the S2ST. Indeed, for the transformation to occur, mechanisms feeding into musical modules, pathways or memory storages would have to become active for interpreting the encoded prosodic contour as musical. Evidence for this proposal comes from the fact that all the prosodic cues identified to facilitate the S2ST contribute to musical (i.e., melodic, rhythmic) contour perception.

As with melody perception, previous research has shown that participants reproduced the intonation contour of a spoken utterance as a melodic contour after having experienced the S2ST. After repeated presentation of the spoken stimulus, pitch reproduction was overall significantly closer to the average original pitch values of the syllables than after a single presentation (Deutsch et al., 2011). Still, slight deviations from these original values seemed to be made in favor of an interpretation as a melodic contour. Together with our results, this indicates that during repetition, a musical pitch class is computed for each or at least several syllables and then integrated in the context of surrounding pitches to form a meaningful contour, whenever possible. In Experiment 1, the presence of stable tonal targets was shown to be particularly helpful during this process. Still, dynamic targets may also undergo pitch class assignment. However, as the present study clearly shows, a high amount of dynamic targets in a sentence was substantially decreasing the chance of experiencing the S2ST. A systematic parametric variation of these pitch features would allow to clarify the issue of how different degrees of cue strength impact on the S2ST in future studies.

The finding that durational contrasts affected the S2ST is particularly intriguing and suggests that mechanisms related to rhythmic contour perception may also contribute to the transformation. In the present study, the S2ST was facilitated by an enhanced durational contrast between accented and unaccented events which was regularly recurring within the sentence (Experiment 2). Importantly, the two rhythmic aspects (i.e., temporal contrast, regular recurrence within the sentence) were perceptually not independent of each other. It is well established that the cognitive ease and pleasure of rhythm processing in a musical context is linked to its fit within a metrical hierarchy (see Keller & Schubert, 2011, for a review). Temporally regular prominence patterns and clear durational ratios between rhythmic events are strong indicators of metrical structure (London, 2004). Therefore, the results from our study indicate that fitting rhythmic patterns of speech to a metrical interpretation may have been facilitating the S2ST. As in the case of pitch, listeners may become more attuned to the temporal relations within a sentence during repetition. Clear durational contrasts may then be more conducive to a metrical interpretation as they underline the prominence structure at local *and* global levels.

Finally, a regular accent structure at a more global level (i.e., across repetitions) did not support the transformation. In contrast, shorter pauses between repetitions facilitated the occurrence of the S2ST. This finding indicates that proximity of repetitions in time constitutes a further aspect affecting the S2ST. Note that the phonological loop in working memory stores incoming phonetic information for a time span of 1 to 2 sec (Baddeley, 1986). This information fades if not reactivated thereafter. It seems plausible that memory for the prosodic contour of our sentences (approx. 2 sec of length) was best reactivated immediately after the end of the sentence in order to stabilize the prosodic percept.

The S2ST is a robust perception phenomenon that is experienced by musicians and nonmusicians alike. However, there are still important individual differences. Some participants experienced the S2ST very often, whereas others never or only occasionally reported the S2ST. Part of this variation can be accounted for by individual differences in sensitivity to rhythmic cues. Rhythmically sensitive participants (Experiment 2) experienced the transformation more often in stimuli with regular accent distribution. In addition to perceptual abilities, it seems possible that more general aspects of decision making may also have played a role during the S2ST. For instance, participants reporting the S2ST very often also reported it earlier than participants with low perception rates. This may indicate that high-perceivers were more confident about their ability to experience the transformation than low-perceivers, or that they were adopting a lower criterion about when a stimulus sounded like sung to them. An unexpected finding was the fact that musicians perceived the transformation generally later than nonmusicians (Experiment 1). Musicians show more distinct encoding and representations of musical, but also of speech prosody (Besson, Chobert & Marie, 2011; Besson, Schön, Moreno, Santos, & Magne, 2007; Dalla Bella, Peretz, & Aronoff, 2003; Lima & Castro, 2011). As the speed of the S2ST is related to the perception of sentence acoustics before repetition, musicians may initially focus more on speech-relevant acoustic properties (e.g., segmental structure, voice timbre, etc.) than nonmusicians, a fact which may delay the transformation. Moreover, musicians have a richer experience with musical stimuli. This may generate a higher criterion for auditory signals to be interpreted as musical. Both issues (perception abilities/strategies; aspects of decision-making during the S2ST) should be addressed in more detail in future studies.

Other massed repetition effects have shown that repetition affects semantic representations as well (Bashford, Warren, & Lenz, 2006, 2008; MacKay et al., 1993; Natsoulas, 1965, 1967; Pilotti, Simcox, Baldy, & Schauss, 2011). Future research could fruitfully exploit the question about potential modifications of the lexico-semantic representation of sentences during the S2ST, given the fact that parallels can be found between the S2ST and semantic satiation. First, both the S2ST and semantic satiation require at least 3 to 4 repetitions to arise (cf. Pilotti et al., 1997). Second, even a slight change of the stimulus' acoustics or its segmental composition blocks the repetition effect in both cases (Deutsch et al., 2011; Pilotti et al., 1997). Third, semantic satiation has been described as a process that induces a reinterpretation of prosodic form (Kuhl & Anderson, 2011). These similarities may point toward some common underlying mechanism across these perceptual transformations.

Why does repetition make such a difference in perception? Repetition has quite different functions in language and music. In speech, immediate repetition of words and sentences is often associated with situations of communicative effort and the aim of clarifying the meaning of an utterance, either because of previous miscommunication, involuntary disfluencies, or rhetorical emphasis (Aitchison, 1994; Bazzanella, 2011). In contrast, musical repetition has been described as being fundamental to the listening experience, being in itself pleasurable and emotionally rewarding (Juslin & Västfjäll, 2008; Livingstone, Palmer, & Schubert, 2012; Margulis, 2013). Furthermore, although one or two subsequent repetitions can occur in speech, multiple repetitions are characteristic of music and structuring elements of some musical forms (e.g., fugue). This functional difference could prime listeners to voluntarily adopt a musical interpretation when excessive repetition hinders its communicative function in speech. In addition, longer utterances seem to be necessary to elicit the S2ST. As demonstrated in our study, a reinterpretation is only possible when a prosodic shape that allows for musical contour perception is present in the stimulus. This may be the reason why single-word repetition effects have not been reported to induce musical perception. Finally, in natural communication, the boundaries between speech and song are not always clear-cut which might also predispose for the S2ST. In early mother–infant communication, for instance, we find a high amount of repetitions and the acoustics of mothers' productions are reminiscent of both speech and song characteristics (e.g., Falk, 2009, 2011; Fernald et al., 1989). Other human communicative activities such as chanting, rapping, ritual speech, or reciting of poetry also naturally fluctuate between speaking and singing, thereby possibly enhancing cognitive fluency and aesthetic pleasure (e.g., Large & Murry, 1980; Reber, Schwarz & Winkielman, 2004).

To conclude, the present study highlights the importance of the S2ST as a general perception phenomenon crossing the boundaries between language and music. The transformation generalizes across languages and is not restricted to skilled audiences (i.e., musicians). We demonstrated that the transformation depends on specific prosodic properties facilitating the formation of a musical percept from a repeated speech signal. This makes this intriguing perceptual phenomenon potentially useful for examining the relations and mutual influences between music and language in terms of shared cognitive resources or mechanisms (Gordon, Schön, Magne, Astésano, & Besson, 2010; Patel, 2008; Schön et al., 2010).

References

Adsett, D. R., & Marchand, Y. (2010). Syllabic complexity: A computational evaluation of nine European languages. *Journal of Quantitative Linguistics, 17,* 269–290. doi:10.1080/09296174.2010.512161

Aitchison, J. (1994). "Say, say it again Sam": The treatment of repetition in Linguistics. In A. Fischer (Ed.), *Repetition* (pp. 15–34). Tübingen, Germany: Gunter Narr.

Arvaniti, A. (2009). Rhythm, timing and the timing of rhythm. *Phonetica, 66,* 46–63. doi:10.1159/000208930

Aschersleben, G. (2002). Temporal control of movements in sensorimotor synchronization. *Brain and Cognition, 48,* 66–79. doi:10.1006/brcg.2001.1304

Auer, P., Couper-Kuhlen, E., & Müller, F. (1999). *Language in time. The rhythm and tempo of spoken interaction*. Oxford, UK: Oxford University Press.

Baddeley, A. D. (1986). *Working memory*. Oxford, UK: Oxford University Press.

Barnes, J., Veilleux, N., Brugos, A., & Shattuck-Hufnagel, S. (2010). Turning points, tonal targets, and the English L- phrase accent. *Language and Cognitive Processes, 25,* 982–1023. doi:10.1080/01690961003599954

Bashford, J. A., Jr., Warren, R. M., & Lenz, P. W. (2006). Polling the effective neighborhood of spoken words with the verbal transformation effect. *Journal of the Acoustical Society of America Express Letters, 119,* EL55–EL59.

Bashford, J. A., Jr., Warren, R. M., & Lenz, P. W. (2008). Evoking biphone neighborhoods with verbal transformations: Illusory changes demonstrate both lexical competition and inhibition. *Journal of the Acoustical Society of America Express Letters, 123,* EL32–EL38.

Bazzanella, C. (2011). Redundancy, repetition, and intensity in discourse. *Language Sciences, 33,* 243–254. doi:10.1016/j.langsci.2010.10.002

Beckman, M., & Edwards, J. (1990). Lengthening and shortening and the nature of prosodic constituency. In J. Kingston & M. Beckman (Eds.), *Papers in laboratory phonology I* (pp. 152–178). Cambridge, UK: University Press.

Besson, M., Chobert, J., & Marie, C. (2011). Language and music in the musician brain. *Language and Linguistics Compass, 5,* 617–634. doi:10.1111/j.1749-818X.2011.00302.x

Besson, M., & Schön, D. (2001). Comparison between language and music. *Annals of the New York Academy of Sciences, 930,* 232–258. doi:10.1111/j.1749-6632.2001.tb05736.x

Besson, M., Schön, D., Moreno, S., Santos, A., & Magne, C. (2007). Influence of musical expertise and musical training on pitch processing in music and language. *Restorative Neurology and Neuroscience, 25,* 399–410.

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International, 5,* 341–345.

Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound*. Cambridge, MA: Bradford Books, MIT Press.

Callan, D., Tsytsarev, V., Hanakawa, T., Callan, A., Katsuhara, M., & Fukuyama, H., & Turner, R. (2006). Song and speech: Brain regions involved with perception and covert production. *NeuroImage, 31,* 1327–1342. doi:10.1016/j.neuroimage.2006.01.036

Cattaneo, Z., Devlin, J. T., Vecchi, T., & Silvanto, J. (2009). Dissociable neural representations of grammatical gender in Broca's area investigated by the combination of satiation and TMS. *NeuroImage, 47,* 700–704. doi:10.1016/j.neuroimage.2009.04.097

Chandrasekaran, B., & Kraus, N. (2010). The scalp-recorded brainstem response to speech: Neural origins and plasticity. *Psychophysiology, 47,* 236–246. doi:10.1111/j.1469-8986.2009.00928.x

Clarke, E. F. (1999). Rhythm and timing in music. In D. Deutsch (Ed.), *The psychology of music,* (pp. 473–500). New York, NY: Academic Press. doi:10.1016/B978-012213564-4/50014-7

Cooper, G., & Meyer, L. B. (1960). *The rhythmic structure of music*. Chicago, IL: University of Chicago Press.

Cummins, F. (2009). Rhythm as an affordance for the entrainment of movement. *Phonetica, 66,* 15–28. doi:10.1159/000208928

Cummins, F., & Port, R. F. (1998). Rhythmic constraints on stress timing in English. *Journal of Phonetics, 26,* 145–171. doi:10.1006/jpho.1998.0070

Cutler, A. (1991). Linguistic rhythm and speech segmentation. In J. Sundberg, J. N. Lennart, & R. Carlson (Eds.), *Music, language, speech and brain* (pp. 157–166). Houndsmills and London, UK: Macmillan.

Dalla Bella, S., & Berkowska, M. (2009). Singing proficiency in the majority: Normality and "phenotypes" of poor singing. *Annals of the New York Academy of Sciences, 1169,* 99–107. doi:10.1111/j.1749-6632.2009.04558.x

Dalla Bella, S., Berkowska, M., & Sowiński, J. (2011). Disorders of pitch production in tone deafness. *Frontiers in Psychology, 2,* 164.

Dalla Bella, S., Białuńska, A., Sowiński, J. (2013). Why movement is captured by music, but less by speech: Role of temporal regularity. *PLoS ONE, 8,* e71945. doi:10.1371/journal.pone.0071945

Dalla Bella, S., Giguère, J.-F., & Peretz, I. (2007). Singing proficiency in the general population. *Journal of the Acoustical Society of America, 121,* 1182–1189. doi:10.1121/1.2427111

Dalla Bella, S., Peretz, I., & Aronoff, N. (2003). Time course of melody recognition: A gating paradigm study. *Perception & Psychophysics, 65,* 1019–1028. doi:10.3758/BF03194831

Dauer, R. (1983). Stress timing and syllable timing reanalyzed. *Journal of Phonetics, 11,* 51–62.

Delaney, P. F., & Verkoeijen, P. P. (2009). Rehearsal strategies can enlarge or diminish the spacing effect: Pure versus mixed lists and encoding strategy. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35,* 1148–1161. doi:10.1037/a0016380

Deutsch, D. (2003). *Phantom words, and other curiosities*. La Jolla, CA: Philomel Records.

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

Drake, C., & Palmer, C. (1993). Accent structures in music performance. *Music Perception, 10,* 343–378. doi:10.2307/40285574

Falk, S. (2009). *Musik und Sprachprosodie. Kindgerichtetes Singen im frühen Spracherwerb*. Berlin, Germany: De Gruyter.

Falk, S. (2011). Melodic vs. intonational coding of communicative functions - A comparison of tonal contours in infant-directed song and speech. *Psychomusicology, 21,* 54–68. doi:10.1037/h0094004

Fernald, A., Taeschner, T., Dunn, J., Papoušek, M., de Boysson-Bardies, B., & Fukui, I. (1989). A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *Journal of Child Language, 16,* 477–501. doi:10.1017/S0305000900010679

Forster, K. I., & Forster, J. (2003). DMDX: A windows display program with millisecond accuracy. *Behavior Research Methods, Instruments & Computers, 35,* 116–124. doi:10.3758/BF03195503

Fox, R., & Qi, Y. (1990). Contextual effects in the perception of lexical tone. *Journal of Chinese Linguistics, 18,* 261–283.

Gillmeister, H., Catmur, C., Liepelt, R., Brass, M., & Heyes, C. (2008). Experience-based priming of body parts: A study of action imitation. *Brain Research, 1217,* 157–170. doi:10.1016/j.brainres.2007.12.076

Goldstein, L. M., & Lackner, J. R. (1972). Alterations of the phonetic coding of speech sounds during repetition. *Cognition, 2,* 279–297. doi:10.1016/0010-0277(72)90036-4

Gordon, R. L., Schön, D., Magne, C., Astésano, C., & Besson, M. (2010). Words and melody are intertwined in perception of sung words: EEG and behavioral evidence. *PLoS One, 5,* e9889. doi:10.1371/journal.pone.0009889

Grabe, E. (1998). Pitch accent realizations in English and German. *Journal of Phonetics, 26,* 129–143. doi:10.1006/jpho.1997.0072

Groppe, D. M., Choi, M., Huang, T., Schilz, J., Topkins, B., Urbach, T. P., & Kutas, M. (2010). The phonemic restoration effect reveals pre-N400 effect of supportive sentence context in speech perception. *Brain Research, 1361,* 54–66. doi:10.1016/j.brainres.2010.09.003

Gussenhoven, C., & Rietveld, T. (1998). On the speaker-dependence of the perceived prominence of F0 peaks. *Journal of Phonetics, 26,* 371–380. doi:10.1006/jpho.1998.0080

Gussenhoven, C., & Rietveld, T. (2000). The behavior of H^* and L^* under variations in pitch range in Dutch rising contours. *Language and Speech, 43,* 183–203. doi:10.1177/00238309000430020301

Harrington, J., Kleber, F., & Reubold, U. (2008). Compensation for coarticulation, /u/-fronting, and sound change in Standard Southern British: An acoustic and perceptual study. *Journal of the Acoustical Society of America, 123,* 2825–2835. doi:10.1121/1.2897042

Helenius, P., Uutela, K., & Hari, R. (1999). Auditory stream segregation in dyslexic adults. *Brain, 122,* 907–913. doi:10.1093/brain/122.5.907

Hillenbrand, J. M., Clark, M. J., & Baer, C. A. (2011). Perception of sinewave vowels. *Journal of the Acoustical Society of America, 129,* 3991–4000. doi:10.1121/1.3573980

Horner, A. J., & Henson, R. N. (2008). Priming, response learning and repetition suppression. *Neuropsychologia, 46,* 1979–1991. doi:10.1016/j.neuropsychologia.2008.01.018

Juslin, P. N., Västfjäll, D. (2008). Emotional responses to music: The need to consider underlying mechanisms. *Behavioral and Brain Sciences, 31,* 559–575. doi:10.1017/S0140525X08005293

Kaminska, Z., & Mayer, P. (2002). Changing words and changing sounds: A change of tune for verbal transformation theory? *European Journal of Cognitive Psychology, 14,* 315–333. doi:10.1080/09541440143000087

Keller, P. E., & Schubert, E. (2011). Cognitive and affective judgements of syncopated musical themes. *Advances in Cognitive Psychology, 7,* 142–156. doi:10.2478/v10053-008-0094-0

Koelsch, S. (2012). *Brain and music*. Hoboken, NJ: Wiley-Blackwell.

Koelsch, S., & Siebel, W. A. (2005). Towards a neural basis of music perception. *Trends in Cognitive Sciences, 9,* 578–584. doi:10.1016/j.tics.2005.10.001

Kondo, H. M., & Kashino, M. (2007). Neural mechanisms of auditory awareness underlying verbal transformations. *NeuroImage, 36,* 123–130. doi:10.1016/j.neuroimage.2007.02.024

Kounios, J., Kotz, S. A., & Holcomb, P. J. (2000). On the locus of the semantic satiation effect: Evidence from event-related brain potentials. *Memory & Cognition, 28,* 1366–1377. doi:10.3758/BF03211837

Krumhansl, C. L. (1990). *Cognitive foundations of musical pitch*. New York, NY: Oxford University Press.

Krumhansl, C. L. (2000). Rhythm and pitch in music cognition. *Psychological Bulletin, 126,* 159–179. doi:10.1037/0033-2909.126.1.159

Krumhansl, C. L. (2005). Musikalische Bezugssysteme. In T. H. Stoffer & R. Oerter (Eds.), *Allgemeine Musikpsychologie* (pp. 267–306). Berlin, Germany: De Gruyter.

Kuhl, B. A., & Anderson, C. A. (2011). More is not always better: Paradoxical effects of repetition on semantic accessibility. *Psychonomic Bulletin & Review, 18,* 964–972. doi:10.3758/s13423-011-0110-0

Kujala, J., Vartiainen, J., Laaksonen, H., & Salmelin, R. (2012). Neural interactions at the core of phonological and semantic priming of written words. *Cerebral Cortex, 22,* 2305–2312. doi:10.1093/cercor/bhr307

Ladd, D. R., & Morton, R. (1997). The perception of intonational emphasis: Continuous or categorical? *Journal of Phonetics, 25,* 313–342. doi:10.1006/jpho.1997.0046

Large, J., & Murry, T. (1980). Quantitative analysis of chant in relation to normal phonation and vocal fry. *Folia phoniatrica, 32,* 14–22. doi:10.1159/000264320

Lehiste, I. (1970). *Suprasegmentals*. Cambridge, MA: MIT Press.

Lehiste, I. (1973). Rhythmic units and syntactic units in production and perception. *Journal of the Acoustical Society of America, 54,* 1228–1234. doi:10.1121/1.1914379

Lehrdahl, F., & Jackendoff, R. (1983). *A generative theory of tonal music*. Cambridge, MA: MIT Press.

Lidji, P., Palmer, C., Peretz, I., & Morningstar, M. (2011). Listeners feel the beat: Entrainment to English and French speech rhythms. *Psycho-*

nomic Bulletin & Review, 18, 1035–1041. doi:10.3758/s13423-011-0163-0

Lieberman, M., & Pierrehumbert, J. (1984). Intonational invariance under changes of pitch range and length. In A. Arnoff & R. Oehrle (Eds.), *Language sound structure* (pp. 157–233). Cambridge, MA: MIT Press.

Lima, C., & Castro, S. L. (2011). Speaking to the trained ear: Musical expertise enhances the recognition of emotions in speech prosody. *Emotion, 11,* 1021–1031. doi:10.1037/a0024521

Lindblom, B. (1978). Final lengthening in speech and music. In E. Garding, G. Bruce, & R. Bannert (Eds.), *Nordic prosody. Papers from a symposium* (pp. 85–101). Lund, Sweden: Lund University.

Livingstone, S. R., Palmer, C., & Schubert, E. (2012). Emotional response to musical repetition. *Emotion, 12,* 552–567. doi:10.1037/a0023747

London, J. (2004). *Hearing in time*. Oxford, UK: Oxford University Press. doi:10.1093/acprof:oso/9780195160819.001.0001

MacKay, D. G., Wulf, G., Yin, C., & Abrams, L. (1993). Relations between word perception and production: New theory and data on the verbal transformation effect. *Journal of Memory and Language, 32,* 624–646. doi:10.1006/jmla.1993.1032

Margulis, E. H. (2013). *On repeat: How music plays the mind*. New York, NY: Oxford University Press. doi:10.1093/acprof:oso/9780199990825.001.0001

Mennen, I., Schaeffler, F., & Docherty, G. (2012). Cross-language differences in fundamental frequency range: A comparison of English and German. *Journal of the Acoustical Society of America, 131,* 2249–2260. doi:10.1121/1.3681950

Nakajima, Y., ten Hoopen, G., & van der Wilk, R. (1991). A new illusion of time perception. *Music Perception, 8,* 431–448. doi:10.2307/40285521

Natsoulas, T. A. (1965). A study of the verbal transformation effect. *The American Journal of Psychology, 78,* 257–263. doi:10.2307/1420498

Natsoulas, T. A. (1967). What are perceptual reports about? *Psychological Bulletin, 67,* 249–272. doi:10.1037/h0024320

Niebuhr, O. (2007). The signaling of German rising-falling intonation categories – The interplay of synchronization, shape, and height. *Phonetica, 64,* 174–193. doi:10.1159/000107915

Nolan, F. (2003). *Intonational equivalence: An experimental evaluation of pitch scales*. Proceedings of the 15th International Congress of Phonetic Sciences, Barcelona, Spain.

Nolan, F., & Asu, E. L. (2009). The pairwise variability index and coexisting rhythms in language. *Phonetica, 66,* 64–77. doi:10.1159/000208931

Patel, A. D. (2003). Language, music, syntax and the brain. *Nature Neuroscience, 6,* 674–681. doi:10.1038/nn1082

Patel, A. D. (2008). *Music, language and the brain*. Oxford, UK: University Press.

Patel, A. D. (2011). Why would musical training benefit the neural encoding of speech? The OPERA hypothesis. *Frontiers in Psychology, 2,* 142. doi:10.3389/fpsyg.2011.00142

Patel, A. D. (2012). Language, music, and the brain: A resource-sharing framework. In P. Rebuschat, M. Rohrmeier, J. Hawkins, & I. Cross (Eds.), *Language and music as cognitive systems* (pp. 204–223). Oxford, UK: Oxford University Press.

Patel, A. D., & Peretz, I. (1997). Is music autonomous from language? A neuropsychological appraisal. In I. Deliège & J. A. Sloboda (Eds.), *Perception and cognition of music* (pp. 191–215). Hove, UK: Psychology Press.

Peretz, I. (2012). Music, Language and modularity in action. In P. Rebuschat, M. Rohrmeier, J. Hawkins, & I. Cross (Eds.), *Language and music as cognitive systems* (pp. 254–268). Oxford, UK: Oxford University Press.

Peretz, I., Champod, A. S., & Hyde, K. (2003). Varieties of musical disorders. The Montreal Battery of Evaluation of Amusia. *Annals of the New York Academy of Sciences, 999,* 58–75. doi:10.1196/annals.1284.006

Peretz, I., & Coltheart, M. (2003). Modularity of music processing. *Nature Neuroscience, 6,* 688–691. doi:10.1038/nn1083

Peretz, I., & Zatorre, R. (2005). Brain organization for music processing. *Annual Review of Psychology, 56,* 89–114. doi:10.1146/annurev.psych.56.091103.070225

Pilotti, M., Antrobus, J. S., & Duff, M. (1997). The effect of presemantic acoustic adaptation on semantic 'satiation'. *Memory & Cognition, 25,* 305–312. doi:10.3758/BF03211286

Pilotti, M., & Khurshid, A. (2004). Semantic satiation in young and older adults. *Perceptual and Motor Skills, 98,* 999–1016.

Pilotti, M., Simcox, T., Baldy, J., & Schauss, F. (2011). Are verbal transformations sensitive to age differences and stimulus properties? *The American Journal of Psychology, 124,* 87–97. doi:10.5406/amerjpsyc.124.1.0087

Pitt, M. A., & Shoaf, L. (2002). Linking verbal transformations to their causes. *Journal of Experimental Psychology: Human Perception and Performance, 28,* 150–162. doi:10.1037/0096-1523.28.1.150

Raaijmakers, J. G. W. (2003). Spacing and repetition effects in human memory: Application of the SAM model. *Cognitive Science, 27,* 431–452. doi:10.1207/s15516709cog2703_5

Rathcke, T. (2013). On the neutralizing status of truncation in intonation: A perception study of boundary tones in German and Russian. *Journal of Phonetics, 41,* 172–185. doi:10.1016/j.wocn.2013.01.003

Reber, R., Schwarz, N., & Winkielman, P. (2004). Processing fluency and aesthetic pleasure: Is beauty in the perceiver's processing experience? *Personality and Social Psychology Review, 8,* 364–382. doi:10.1207/s15327957pspr0804_3

Remez, R. E., Rubin, P. E., Pisoni, D. B., & Carrell, T. D. (1981). Speech perception without traditional speech cues. *Science, 212,* 947–949. doi:10.1126/science.7233191

Repp, B. H. (2005). Sensorimotor synchronization: A review of the tapping literature. *Psychonomic Bulletin & Review, 12,* 969–92. doi:10.3758/BF03206433

Repp, B. H. (2006). Musical synchronization. In E. Altenmüller, J. Kesselring, & M. Wiesendanger (Eds.), *Music, motor control, and the brain* (pp. 55–76). Oxford, UK: Oxford University Press.

Roach, P. (1982). On the distinction between stress-timed languages and syllable-timed languages. In D. Crystal (Ed.), *Linguistic controversies: Essays in honour of F. R. Palmer* (pp. 73–79). London, UK: Arnold.

Samuel, A. G. (1981). Phonemic restoration: Insights from a new methodology. *Journal of Experimental Psychology: General, 110,* 474–494. doi:10.1037/0096-3445.110.4.474

Sasaki, T., Suetomi, D., Nakajima, Y., & ten Hoopen, G. (2002). Time shrinking, its propagation and Gestalt principles. *Perception & Psychophysics, 64,* 919–931. doi:10.3758/BF03196796

Sato, M., Schwartz, J. L., Abry, C., Cathiard, M. A., & Loevenbruck, H. (2006). Multistable syllables as enacted percepts: A source of an asymmetric bias in the verbal transformation effect. *Perception & Psychophysics, 68,* 458–474. doi:10.3758/BF03193690

Sato, M., Vallée, N., Schwartz, J.-L., & Rousset, I. (2007). A perceptual correlate of the labial-coronal effect. *Journal of Speech, Language and Hearing Research, 50,* 1466–1480. doi:10.1044/1092-4388(2007/101)

Schellenberg, E. G., & Trehub, S. E. (1996). Natural music intervals. Evidence from infant listeners. *Psychological Science, 7,* 272–277. doi:10.1111/j.1467-9280.1996.tb00373.x

Schön, D., Gordon, R., Campagne, A., Magne, C., Astésano, C., Anton, J. L., & Besson, M. (2010). Similar cerebral networks in language, music and song perception. *NeuroImage, 51,* 450–461. doi:10.1016/j.neuroimage.2010.02.023

Selkirk, E. O. (1981). English compounding and the theory of word-structure. In M. Moortgat, H. van der Hulst, & T. Hoestra (Eds.), *The scope of lexical rules* (pp. 229–277). Dordrecht, The Netherlands: Foris.

Simko, J., & Cummins, F. (2010). Embodied task dynamics. *Psychological Review, 117,* 1229–1246. doi:10.1037/a0020490

Smith, L. C. (1984). Semantic satiation affects category membership decision time but not lexical priming. *Memory & Cognition, 12,* 483–488. doi:10.3758/BF03198310

Souvignier, T. (2003). *Loops and grooves: The musician's guide to groove machines and loop sequencers*. New York, NY: Hal Leonard.

Stewart, L., von Kriegstein, K., Dalla Bella, S., Warren, J. D., & Griffiths, T. D. (2009). Disorders of musical cognition. In S. Hallam, I. Cross, & M. Thaut (Eds.), *Oxford handbook of music psychology* (pp. 184–196), New York, NY: Oxford University Press.

Sundberg, J. (1989). Synthesis of singing by rule. In M. V. Mathews & J. R. Pierce (Eds.), *Current directions in computer music research* (pp. 45–55). Cambridge, MA: MIT Press.

Tan, S. L., Pfordresher, P. Q., & Harré, R. (2010). *Psychology of music: From sound to significance*. London, UK: Routledge and Psychology Press.

t'Hart, J. (1981). Differential sensitivity to pitch distance, particularly in speech. *Journal of the Acoustical Society of America, 69,* 811–821.

t'Hart, J., & Cohen, A. (1973). Intonation by rule: A perceptual quest. *Journal of Phonetics, 1,* 309–327.

t'Hart, J., Collier, R., & Cohen, A. (1990). *A perceptual study of intonation*. Cambridge, UK: Cambridge University Press.

Tierney, A., Dick, F., Deutsch, D., & Sereno, M. (2012). Speech versus song: Multiple pitch-sensitive areas revealed by a naturally occurring musical illusion. *Cerebral Cortex*. doi:10.1093/cercor/bhs003

Tillmann, B., Rusconi, E., Traube, C., Butterworth, B., Umilta, C., & Peretz, I. (2011). Fine-grained pitch processing of music and speech in congenital amusia. *Journal of the Acoustical Society of America, 130,* 4089–4096. doi:10.1121/1.3658447

Turk, A., & Shattuck-Hufnagel, S. (2007). Phrase-final lengthening. *American English Journal of Phonetics, 35,* 445–472. doi:10.1016/j.wocn.2006.12.001

Villing, R. C., Repp, B. H., Ward, T. E., & Timoney, J. M. (2011). Measuring perceptual centers using the phase correction response. *Attention, Perception, & Psychophysics, 73,* 1614–1629. doi:10.3758/s13414-011-0110-1

Ward, G., & Hirschberg, J. (1985). Implicating uncertainty: The pragmatics of fall-rise intonation. *Language, 61,* 747–776. doi:10.2307/414489

Warren, R. M. (1961). Illusory changes of distinct speech upon repetition: The verbal transformation effect. *British Journal of Psychology, 52,* 249–258. doi:10.1111/j.2044-8295.1961.tb00787.x

Warren, R. M. (1968). Verbal transformation effect and auditory perceptual mechanisms. *Psychological Bulletin, 70,* 261–270. doi:10.1037/h0026275

Warren, R. M. (1970). Perceptual restoration of missing speech sounds. *Science, 167,* 392–393. doi:10.1126/science.167.3917.392

Warren, R. M. (2008). *Auditory perception: An analysis and synthesis*. New York, NY: Cambridge University Press. doi:10.1017/CBO9780511754777

Warren, R. M., Bashford, J. A., Jr., & Gardner, D. A. (1990). Tweaking the lexicon: Organization of vowel sequences into words. *Perception & Psychophysics, 47,* 423–432. doi:10.3758/BF03208175

Warren, R. M., & Gregory, R. L. (1958). An auditory analogue of the visual reversible figure. *The American Journal of Psychology, 71,* 612–613. doi:10.2307/1420267

Warren, R. M., Healy, E. W., & Chalikia, M. H. (1996). The vowel-sequence illusion: Intrasubject stability and intersubject agreement of syllabic forms. *Journal of the Acoustical Society of America, 100,* 2452–2461. doi:10.1121/1.417953

Warren, R. M., & Meyers, M. D. (1987). Effects of listening to repeated syllables: Category boundary shifts versus verbal transformations. *Journal of Phonetics, 15,* 169–181.

Xu, Y. (1994). Production and perception of coarticulated tones. *Journal of the Acoustical Society of America, 95,* 2240–2253. doi:10.1121/1.408684

Xue, G., Mei, L., Chen, C., Lu, Z., Poldrack, R. A., & Dong, Q. (2011). Spaced learning enhances subsequent recognition memory by reducing neural repetition suppression. *Journal of Cognitive Neuroscience, 23,* 1624–1633. doi:10.1162/jocn.2010.21532

Zago, L., Fenske, M. J., Aminoff, E., & Bar, M. (2005). The rise and fall of priming: How visual exposure shapes cortical representations of objects. *Cerebral Cortex, 15,* 1655–1665. doi:10.1093/cercor/bhi060

Zatorre, R. J., & Baum, S. R. (2012). Musical melody and speech intonation: Singing a different tune. *PLoS Biology, 10,* e1001372. doi:10.1371/journal.pbio.1001372

Zatorre, R. J., Belin, P., & Penhune, V. B. (2002). Structure and function of auditory cortex: Music and speech. *Trends in Cognitive Sciences, 6,* 37–46. doi:10.1016/S1364-6613(00)01816-7

Zatorre, R. J., & Gandour, J. T. (2008). Neural specializations for speech and pitch: Moving beyond the dichotomies. *Philosophical Transactions of the Royal Society B, 363,* 1087–1104. doi:10.1098/rstb.2007.2161

Received September 7, 2012
Revision received April 3, 2014
Accepted April 7, 2014 ■