BRIEF REPORT

CrossMark

There is music in repetition: Looped segments of speech and nonspeech induce the perception of music in a time-dependent manner

Jess Rowland [1,2] · Anna Kasdan [1,3] · David Poeppel [1,4]

Abstract

While many techniques are known to music creators, the technique of repetition is one of the most commonly deployed. The mechanism by which repetition is effective as a music-making tool, however, is unknown. Building on the speech-to-song illusion (Deutsch, Henthorn, & Lapidis in *Journal of the Acoustical Society of America*, *129*(4), 2245–2252, 2011), we explore a phenomenon in which perception of musical attributes are elicited from repeated, or "looped," auditory material usually perceived as nonmusical such as speech and environmental sounds. We assessed whether this effect holds true for speech stimuli of different lengths; nonspeech sounds (water dripping); and speech signals decomposed into their rhythmic and spectral components. Participants listened to looped stimuli (from 700 to 4,000 ms) and provided continuous as well as discrete perceptual ratings. We show that the regularizing effect of repetition generalizes to nonspeech auditory material and is strongest for shorter clip lengths in the speech and environmental cases. We also find that deconstructed pitch and rhythmic speech components independently elicit a regularizing effect, though the effect across segment duration is different than that for intact speech and environmental sounds. Taken together, these experiments suggest repetition may invoke active internal mechanisms that bias perception toward musical structure.

Keywords Speech-to-song illusion · Repetition effects · Music cognition · Sound recognition

Repetition is a common method employed by music creators (Margulis, 2014), and it affords us musical beat, melodic motifs, and rhythms (Handel, 1989; London, 2002). At larger scales, we experience melodic integration, where repetition's role becomes more nuanced and structural (Narmour, 1992; Ockelford, 2005; Rahn, 1993). The ubiquity of temporal regularity across multiple time scales suggests auditory repetition engages a mechanism central to our experience of "musicality."

✉ Jess Rowland
   antidada@gmail.com

1   Department of Psychology, New York University, 6 Washington Place, Room 275, New York, NY 10003, USA

2   Lewis Center for the Arts, Princeton University, Princeton, NJ 08544, USA

3   Vanderbilt Brain Institute, Vanderbilt University Medical Center, Nashville, TN 37232, USA

4   Max Planck Institute for Empirical Aesthetics, 60322 Frankfurt, Germany

While these repetition effects of musical material are well known, repetition of auditory material not normally perceived as musical, such as speech and environmental sounds (Patel, 2003), can also generate musical perception. This effect was compellingly described by Deutsch, Henthorn, and Lapidis (2011), who discovered that repeated speech can perceptually transform into perceived singing, an effect known as the speech-to-song illusion. In their initial study, when a single, unchanging spoken phrase was repeated and subjects were asked to rate how much it sounded like speech or singing, a significant increase in perceived singing was found after repetition. This phenomenon suggests repetition can distort the perception of speech from prosody toward pitched music.

Follow-up studies of the speech-to-music illusion support the idea that the ability to focus on tonal properties of stimuli increases the size of this effect. Falk, Rathcke, and Dalla Bella (2014) showed that a strong predictor of the effect is strongly pitched prosodic cues, and Margulis et al. (2015) found that the illusion is induced more quickly when using nonnative languages and languages difficult to pronounce relative to native language. More recently, Graber, Simchy-Gross, and Margulis (2017) showed that subjects had reduced sensitivity to absolute

pitch manipulations after speech-to-song transformation, suggesting this illusion affects pitch perception. The neural basis of the speech-to-song illusion has also been investigated in fMRI studies showing that anterior temporal activity associated with pitch processing increases as the perceptual effect increases (Tierney, Dick, Deutsch, & Sereno, 2013).

Most recently, Simchy-Gross and Margulis (2018) used the original Deutsch et al. (2011) paradigm with a set of environmental sounds to show that repetition-induced musical perception extends to nonspeech sounds. Composers are also aware of these effects for repeated nonmusical material—and not just for speech. Repetition of nonmusical material is a common compositional strategy in musique concrète (Schaeffer, 1952), minimalism (Fink, 2005), and new music (Bosetti, 2012).

If this phenomenon is representative of a general mechanism, it should extend beyond speech and may exist for spectral and rhythmic components of speech independently. Such a phenomenon might also predominate under preferred temporal conditions if it were tied to auditory processing mechanisms well-known to be sensitive to temporal structure (Arnal et al., 2015a, b; Zatorre et al., 2002).

We tested the extent to which repetition's increased perception of musical attributes extends to (1) speech clips of varying lengths; (2) environmental sounds; deconstructed speech with only (3) rhythmic content and (4) spectral content. We replicate and extend this effect of repetition to these categories. Furthermore, the effect shows a preferred temporal range: Shorter-duration stimuli in speech and environmental sounds, but not deconstructed stimuli, increase the effect.

Experiments 1 and 2: Generalizability of the perceptual effect

Method

Participants Thirty participants took part in Experiment 1 (mean age = 19 years; age range: 18–22; 28 female) and 30 participants took part in Experiment 2 (mean age = 20 years; age range: 18–28; 24 female). There were no overlapping participants between Experiments 1 and 2. The number of participants was set to be consistent with previous work on the speech-to-song illusion (Falk et al., 2014; Margulis & Simchy-Gross, 2016; Margulis, Simchy-Gross, & Black, 2015), and especially to the original Deutsch et al. (2011) finding for direct comparison. All experiments were conducted with procedures approved by the NYU Committee on Activities Involving Human Participants.

Materials For all experiments, no stimuli appeared in more than a single trial. In both Experiments 1 and 2, 42 stimuli (trials) were randomly interleaved. Trials were binned for analysis (see Fig. 1b) as follows: We used six bins divided

by stimuli length, with seven stimuli per bin. Bins were set such that average stimuli length for each bin ranged from 1.0 – to 3.5 s in length, with each successive bin average length separated by 0.5 s. Both Experiments 1 and 2 contained the same number of bins and same mean stimulus length per bin. Experiments 3 and 4 used acoustically modified clips from Experiment 1, and therefore matched binning of Experiments 1 and 2.

Experiment 1: Speech stimuli All speech stimuli were recordings of grammatically correct spoken English phrases or sentences, spanning a length range from 700 ms to 4,000 ms. All speech audio clips were public domain recordings from Libravox (https://librivox.org/), taken from one female speaker, reciting relatively obscure texts (to avoid any possible familiarity effects) by Arthur Chesterton, Fanny Coe, and an anonymous author of *The Broken Vase*.

Speech clips were checked for noise and acoustic artifacts, and RMS was normalized to 70 dB using Praat. A linear 30-ms fade in and fade-out was implemented in Audacity to prevent effects of transients ("clicks") in the audio signal at the clips' beginning and end.

- ***Experiment 2: Environmental (water) sounds.*** All environmental sound clips were unique clips taken at random from one single recording of nonisochronous dripping water recorded on a Zoom H4n field recorder, spanning a length range from 700 ms to 4,450 ms. All clips were checked for noise and artifacts, and peak normalized (−1.0 dB) using Audacity.

Dripping water is a common environmental sound and has been used previously effectively to study nonspeech auditory material (e.g., Simchy-Gross & Margulis, 2018). Water-like sound textures have also been used in recent work on acoustic classification (e.g., McDermott, Schemitsch, & Simoncelli, 2013.)

Procedure The procedure for Experiment 1 and Experiment 2 was identical, except different categories of sound were tested: in Experiment 1, speech stimuli, and Experiment 2, environmental sounds. The experimental design is outlined in Fig. 1a.

In both experiments, participants sat at a presentation computer and listened to stimuli presented monophonically over Sennheiser HD 380 headphones. A given trial had three components, sequentially: (1) prerepetition discrete judgment, (2) continuous judgment, and (3) postrepetition discrete judgment.

- ***Prerepetition discrete judgment***. Participants were first presented with a single exposure of a given stimulus and asked to make a self-paced judgment about the stimulus's musicality. In Experiment 1, the musicality question was: "How much did this sound like speech or singing?",
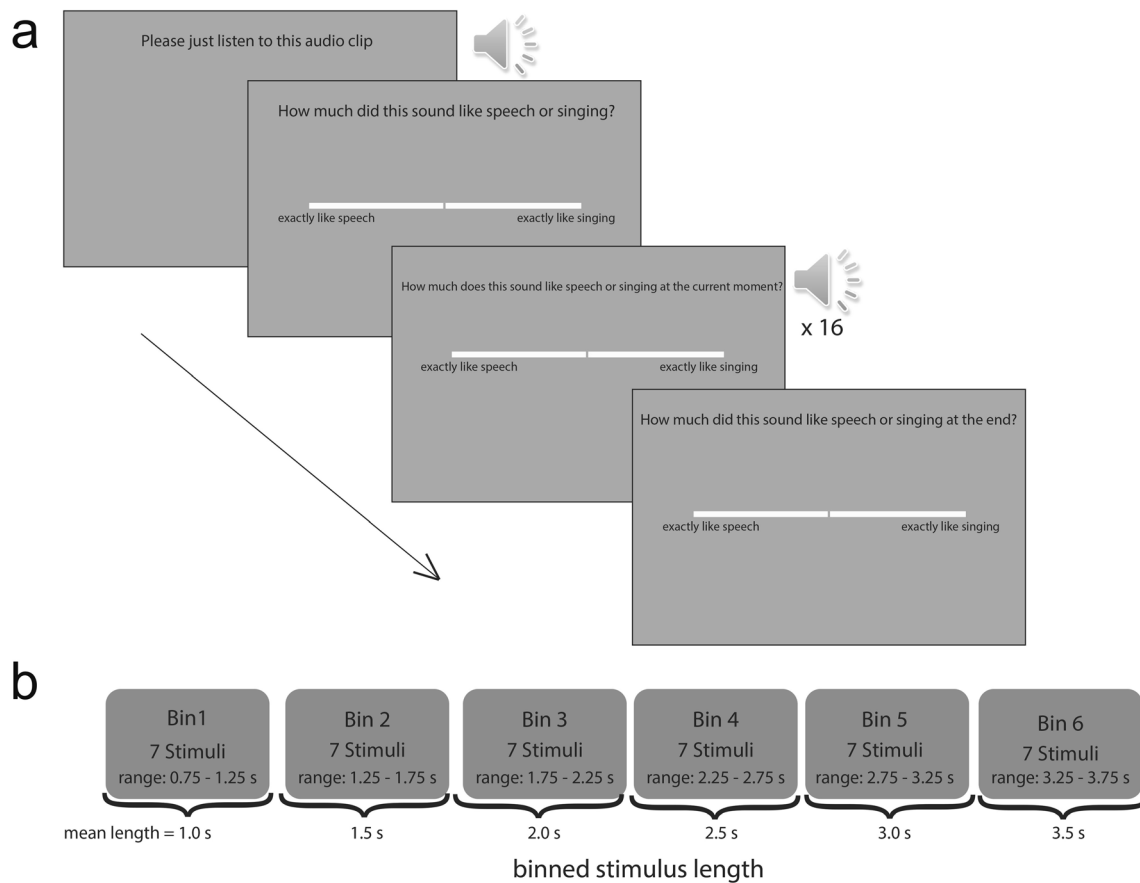
Fig. 1 **a** Trial structure. Participants were instructed to listen to a single presentation of an audio clip, after which they made a discrete prerepetition judgement of the clip's musicality using a mouse. Next, the trial stimulus was looped 16 times while participants used the same method for continuous online rating. At the end of the looped presentation, participants made a second, postrepetition, discrete judgement of the same stimuli. **b** Binning procedure. Stimuli for all experiments were divided into six different bin lengths, with stimuli in each bin spanning a 0.5-s range. There were seven stimuli in each of the six bins, for a total of 42 stimuli used in both Experiments 1 and 2

matching the procedure of Deutsch et al. (2011) to directly compare a wide, parameterized set of speech clips with the original single-exemplar finding of Deutsch et al. (2011). In Experiment 2, participants were asked, "How much did this sound like music?"

As the question was displayed, a vertical slider bar appeared on-screen, and participants were asked to denote their judgment. For Experiment 1, the slider bars' left end was marked "Exactly like speech" and the right was marked "Exactly like singing." For Experiment 2, the left end was marked "Not at all" and the right "Very much."

Participants then moved the mouse until determining that the mouse position accurately reflected their judgment, at which point they clicked the mouse to register this discrete response. The initial slider position was set to the middle of the bar.

- **Continuous judgment.** Participants provided continuous ratings so their-real time judgment of perceived musicality during repetition could be measured. Continuous ratings have been employed in studies of music listening to assess participants' judgments of musical features such as emotional content, musical themes, and tension (Farbood, 2016; Mas-Herrero, Zatorre, Rodriguez-Fornells, & Marco-Pallarés, 2014; Wen & Krumhansl, 2017).

The given stimulus was then repeated 16 times without pause. During repetition, participants made continuous judgements, sampled at 60 Hz, along the vertical slider bar by sliding the mouse. The initial slider position corresponded with the initial judgement of the first discreet presentation.

- **Postrepetition discrete judgment.** At the conclusion of repeated auditory stimuli, participants were asked to make a second discrete, self-paced postrepetition judgement, using the same methodology as the first discreet judgement. In this case, the slider position defaulted to the position at the end of the trial's continuous phase.

Before each experiment, participants performed four practice trials. All participants confirmed that they understood task instructions before beginning the actual experiment.

Results

Exploring continuous ratings Continuous ratings of musicality reflect a diversity of responses. Figure 2a shows randomly selected raw time series from 10 participants to one single stimulus in Experiment 1. The heterogeneity of response profiles suggests that participants have multiple strategies for approaching the continuous rating task. Despite overall differences in absolute ratings and apparent strategies for response, mean traces for all 30 participants in Experiment 1 resolved to a consistent pattern: The shorter the stimuli, by condition, the stronger the effect (see Fig. 2b). The rate of change for the overall mean trace of Experiment 1 showed that the effect proceeds at a continuously positive rate of change, and by 16 repeats there is a slow-down in the rate of change (see Fig. 2c), suggesting a saturation of the effect.

Replicating the speech-to-song effect and extending it to nonspeech stimuli Comparing postrepetition to prerepetition ratings in Experiment 1 using speech stimuli (see Fig. 3a), we found a significant positive effect of repetition on musicality ratings, $t(58) = 5.27$, $p < .001$, $d = 1.36$. This finding replicates the originally reported speech-to-song effect (Deutsch et al., 2011) using a large set of distinct stimuli. However, the size of the effect found here was less than reported in Deutsch et al. (2011). With environmental sounds (Experiment 2), the same effect was found: postrepetition ratings of musicality were significantly higher than prerepetition ratings (see Fig. 2a), $t(58) = 8.9665$, $p < .001$, $d = 2.32$, similar to the effect found by Simchy-Gross and Margulis (2018).

Trials in the two experiments were then broken down by stimulus length (or "loop size") into six bins for both Experiments 1 and 2 (see Fig. 3b). In both experiments, shorter stimulus lengths (or "loop length") created a larger effect, despite participants having less absolute time to respond. The trend of this effect is significant for both speech (rank sum $p < .001$, $z = 6.65$, mean $r = -0.545$; $SD = 0.479$) and environmental sounds (rank sum $p < .001$, $z = 6.65$, mean $r = -0.3517$; $SD = 0.483$). The mean slope for both sets of stimuli trended downward and were not significantly different, $t(58) = -1.56$; $p = .12$, $d = -0.40$.

Experiments 3 and 4: The role of rhythmic and pitch components in repetition-induced musicality

Method

Participants For Experiment 3, 20 participants were tested (mean age = 20 years; age range: 18–24; 15 female). In Experiment 4, a separate set of 20 participants was tested (mean age = 19 years; age range: 18–24; 15 female). There

were no overlapping participants for Experiments 3 and 4. For both experiments, musically trained subjects were tested to ensure understanding of task instructions and accurate reporting of the percept (i.e., they were asked about musical rhythm and "sequence of notes"): Experiment 3, mean training = 10 years, range: 4–20 years; Experiment 4, mean training = 9 years, range: 5–16 years. Previous results have shown that both musically trained and untrained subjects experience the speech-to-song effect (Falk et al., 2014; Vanden Bosch der Nederlanden, Hannon, & Snyder, 2015). All experiments were conducted in accordance with procedures approved by the NYU Committee on Activities Involving Human Participants.

Materials The set of speech stimuli used in Experiment 1 was deconstructed into their rhythmic (Experiment 3) and pitch (Experiment 4) components as follows:

- **Experiment 3: Rhythm clips.** Rhythm clips were created to test the rhythmic component of speech stimuli with minimized spectral variation. Each rhythm clip contained a sequence of percussive cross-stick samples (i.e., two drumsticks striking together) corresponding to each clip's rhythmic content. The samples' spectral content did not vary from one event to another.

The placement of the cross-stick samples was determined in a semiautomated fashion using the MATLAB-based Music Information Retrieval Toolbox (Lartillot & Toiviainen, 2007). Onsets and energy peaks of syllables, which are known to correspond to perceived speech rhythm (Ding et al., 2017; Ghitza, 2013) were determined by peak detection using the mironsets function of the Music Information Retrieval Toolbox (Lartillot & Toiviainen, 2007).

A new rhythm clip was generated by placing concatenated cross-stick samples at event onsets. These rhythm clips were manually checked by an expert operator against original speech clips to confirm that the algorithm generated perceptually accurate rhythms corresponding to the speech rhythms. In cases where the algorithm performed poorly (about one-third of cases), manual adjustments were made using Adobe Audition. All rhythm clips were peak normalized with 30 ms linear fade-in and fade-out. Rhythm clips were then binned according to the procedure in Experiment 1.

For control stimuli, the same procedure for generating rhythmic stimuli was used on nonrepeating 30-s speech excerpts, rather than multiple loops of smaller excerpts. Seven unique control trial stimuli were each presented once during the experiment matching the seven clips for each "loop"-size bin.

- **Experiment 4: Pitch clips.** Temporally filtered clips were created from the original speech stimuli of Experiment 1 to test whether the pitch component of the speech stimuli (without significant rhythmic
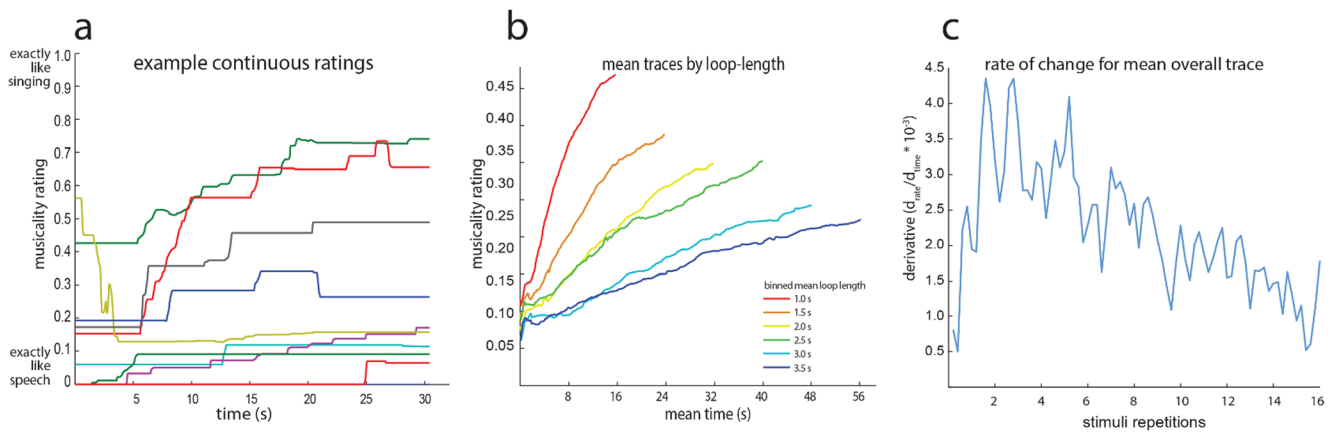
Fig. 2 Time course of the speech-to-music effect. **a** Representative sample of participants' raw musicality ratings over time in Experiment 1, from the start to the end of looped repetition for one specific stimulus. **b** Mean traces separated into binned "loop" size from the shortest (1 s average loop size × 16 repeats) to longest (3.5 s average loop size × 16 repeats) for all participants. **c** Rate of change for the overall mean trace by number of repetitions. (Color figure online)

content) would elicit a repetition effect. To accomplish this, temporal modulations were removed from the speech clips using a low-pass temporal modulation filter with a 1.5-Hz cutoff (see Method; Arnal et al., 2015a, b), as follows: Each original speech clip's time frequency representation was spectrally decomposed using a 128-channel filter bank and transformed into the modulation domain and low-pass filtered at 1.5 Hz. The modulation-domain information was then inverted back to the time domain through iterative convex projection. This retained only the most basic level of time-varying

spectral content, while fully removing the rhythmic components of spoken speech that exist primarily within the theta range (4–7 Hz) (Ding et al., 2017).

For control stimuli, the same procedure for generating temporally modulated stimuli was used on nonrepeating 30-s speech excerpts, rather than multiple loops of smaller excerpts. Seven unique control-trial stimuli were each presented once during the experiment matching the seven clips for each "loop"-size bin. All clips were RMS normalized, with linear fade-in and fade-out of 30 ms.
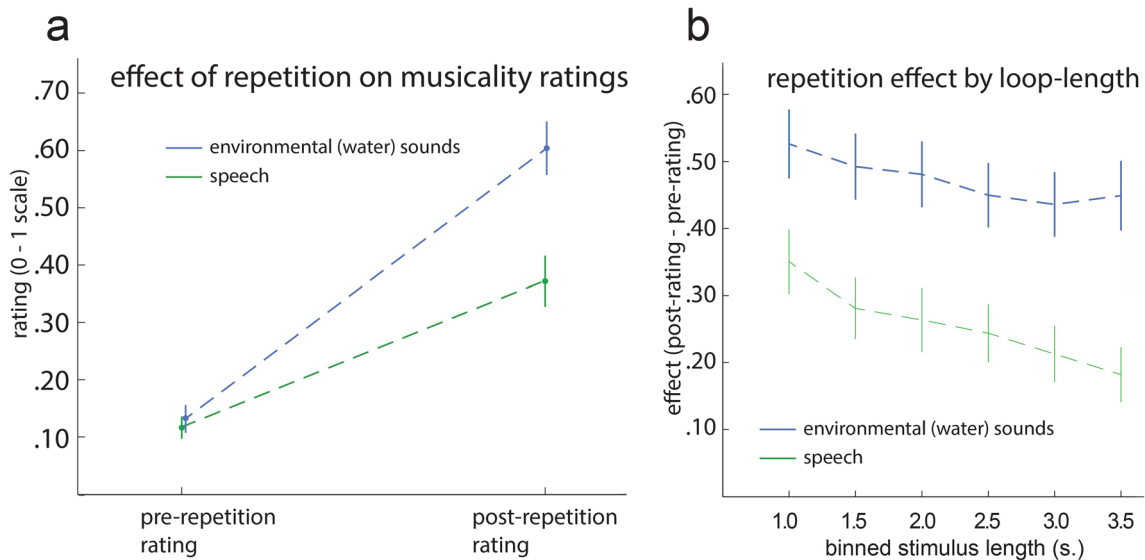
Fig. 3 Confirming and extending the speech-to-music effect to nonspeech sounds. **a** Comparison of discrete prerepetition judgements to postrepetition judgements on a 0–1 scale showing repetition-induced effects for both speech (Experiment 1) and nonspeech (Experiment 2,) sounds. **b** Difference between postrepetition and prerepetition judgments separated into binned loop lengths, showing a downward trend from shortest stimulus length to the longest. Bars are standard error for both plots. Since the context-dependent wording of the musicality judgment question varied between the two experiments, only within-experiment effect-size comparisons—but not comparison between the two experiments—are warranted

Procedure The experimental procedures for Experiments 3 and 4 were identical to Experiment 1 (see Fig. 1), except with modified musicality judgment questions in each case. In Experiment 3 (rhythm clips), the question was: "How much did this pattern sound like it had a musical beat to you?" In Experiment 4 (pitch clips), the questions was: "How much did this sound like a sequence of notes to you?" In both experiments, the left end of the scale was *none at all* and the right end was *very much*.

Results

Comparing postrepetition to prerepetition ratings (see Fig. 4), we found a significant effect of repetition exposure on musicality ratings for rhythm (Experiment 3), $t(38) = 4.59$, $p < .001$, $d = 1.45$, and pitch (Experiment 4), $t(38) = 3.96$, $p < .001$, $d = 1.25$. These results with temporally modified and rhythmic stimuli mirror those found with speech and environmental sounds. Control stimuli did not show any positive change in musicality ratings postexposure for both rhythm (-0.31 mean rating change), $t(38) = -6.25$, $p < .001$, $d = 1.98$, and pitch (-0.24 mean rating change), $t(38) = -3.5603$, $p = .001$, $d = 1.26$, actually decreasing, consistent with an accumulation of evidence against musical attributes.

Trials in both experiments were then broken down by stimulus length (or "loop length") into the same six bins used in Experiment 1 (see Fig. 3b). Visual inspection reveals that the curve's shape in both cases does not exhibit the consistent downward slope seen in Experiments 1 and 2, instead revealing a more uniform dynamic. A one-way ANOVA showed no significance across loop length for either experiment, Experiment 3, $F(5, 1) = 0.79$, $p = .56$; Experiment 4, $F(5, 1) = 0.47$, $p = .80$.

Discussion

We first replicated the speech-to-music effect using a large corpus of speech stimuli parameterized for duration, showing significant increase in perceived musicality after multiple repetitions. Second, we found this effect extends to environmental (water droplet sounds) looped with the same temporal constraints as the speech stimuli. For both experiments, we found that shorter loop lengths elicited the strongest perceptual effect. When we deconstructed the speech signal, we were able to generate perception of musical attributes from (1) the speech rhythm alone and (2) pitch content alone. The apparent ubiquity of repetition-induced perceived musical attributes using different acoustic and environmental categories suggests a general mechanism not specifically tied to speech, or any particular component (spectral or rhythmic) in the signal. The robust illusion described by Deutsch et al. (2011) may be a special case of a broader phenomenon encompassing generalized repeated auditory material, better described as a "repetition-to-music" effect.

This repetition-to-music effect can be described as a related cluster of findings, as each experiment uses a different
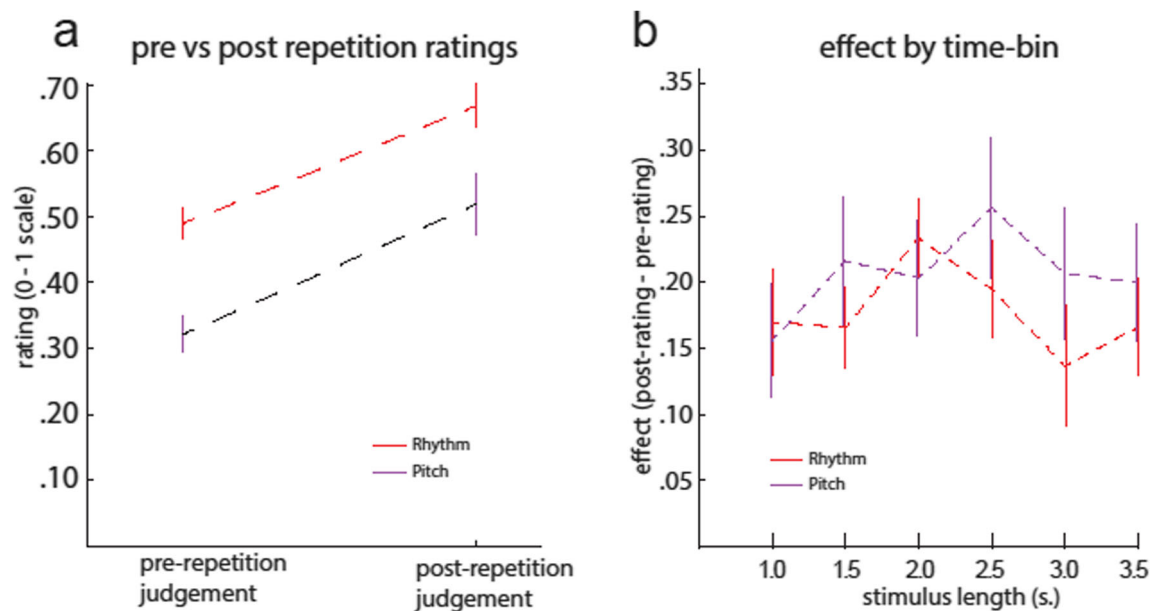
Fig. 4 Role of pitch and rhythm components on repetition-induced musicality. **a** Comparison of discreet prerepetition judgements to postrepetition judgements on a 0–1 scale, showing positive repetition-induced effects for both rhythm (Experiment 3) and pitch (Experiment 4). **b** Difference between postrepetition and prerepetition judgments separated into binned loop lengths. Bars are standard errors for both plots. Note that since the context-dependent wording of the musicality judgment question varied between the two experiments, only within-experiment effect-size comparisons—but not comparison between the two experiments—are warranted

measure based on the context of material being repeated, whether it be the perception of "song," "music," "notes," or "beat." That said, in all cases, repetition consistently drives perception toward musical qualities.

Several researchers (Deutsch et al., 2011; Falk et al., 2014; Tierney et al., 2013) suggest that low-level processes can latch onto tonal properties of repeated speech, and it is this process by which perception of musical attributes appears from repetition. However, this can only be part of the explanation, for two reasons. First, as we show in Experiment 4, the percept following repetition changes to conform closer to a sequence of musical notes rather than toward the stimuli's nonmusically pitched acoustics, as might be expected from improved saliency. Second, our evidence shows that repetition affects rhythm as well as pitch (i.e., both temporal and spectral domains).

In this context, the work of Desain and Honing (2003) is illuminating. They found that repeated presentation of random rhythms (at similar temporal scales used here) is biased toward perception of musical patterns, and suggested this is accomplished by fitting rhythms into categorical classes of rhythmic perception. Such internal musical representations have long been proposed (e.g., Longuet-Higgins & Lee, 1982; Palmer & Krumhansl, 1990) and may be at work in the present results. Further, our finding of a more pronounced effect for shorter loops implicates temporal windows of early auditory system processing (for review, see Haegens and Zion Golumbic, 2017) which are well-known for both speech (Ding et al., 2017; Teng, Tian, & Poeppel, 2016), music (Doelling & Poeppel, 2015), and, specifically, musical rhythm (Large & Snyder 2009; Nozaradan 2014; Tierney & Kraus, 2015).

It may be possible that at shorter clip durations, speech and environmental sound categorization is more tenuous, making stimuli more readily available to be perceived outside these categories. Some evidence exists that, at least for speech, categorization ability can play a role in effect strength (Margulis et al., 2015). While we cannot rule out this possibility, clear speech is known to be readily categorized at durations much shorter than those presented here (e.g., Overath, McDermott, Zarate, & Poeppel, 2015).

The results here support the general mechanism of a repetition-to-music effect and suggest that music can be, though not necessarily *must* be, generated de novo in the mind, from general auditory input. It may be that music creators use this tool of repetition to co-opt internal preexisting mechanisms for musical purposes—an idea that comports with contemporary composers' exploration of the notion that under the right conditions anything can be processed as musical (Cage, 1961).

References

Arnal, L. H., Poeppel, D., & Giraud, A.-L. (2015a). Temporal coding in the auditory cortex. *Handbook of Clinical Neurology, 129*, 85–98.

Arnal, L. H., Flinker, A., Kleinschmidt, A., Giraud, A.-L., & Poeppel, D. (2015b). Human screams occupy a privileged niche in the communication soundscape. *Current Biology, 25*, 1–6.

Bosetti, A. (2012). *Speech music and/or.* Retrieved from http://www.melgun.net/read/speech-music-and-or-alessandro-bosetti-in-conversation-with-peter-ablinger-tomomi-adachi-arturas-bumsteinas-jenny-walshe-and-alexander-waterman/

Cage, J. (1961). Silence: Lectures and writings. Middletown: Wesleyan University Press.

Desain, P., & Honing, H. (2003). The formation of rhythmic categories and metric priming. *Perception, 32*(3), 341–365.

Deutsch, D., Henthorn, T., & Lapidis, R. (2011). Illusory transformation from speech to song. *The Journal of the Acoustical Society of America, 129*(4), 2245–2252.

Ding, N., Patel, A. D., Chen, L., Butler, H., Luo, C., & Poeppel, D. (2017). Temporal modulations in speech and music. *Neuroscience & Biobehavioral Reviews, 81*(Part B), 181–187. https://doi.org/10.1016/j.neubiorev.2017.02.011.

Doelling, K. B., & Poeppel, D. (2015). Cortical entrainment to music and its modulation by expertise. *Proceedings of the National Academy of Sciences of the United States of America, 112*(45), E6233–E6242. doi:https://doi.org/10.1073/pnas.1508431112

Falk, S., Rathcke, T., & Dalla Bella, S. (2014). When speech sounds like music. *Journal of Experimental Psychology: Human Perception and Performance, 40*(4), 1491–1506.

Farbood, M. M. (2016). Memory of a tonal center after modulation. *Music Perception, 34*(1), 71–93.

Fink, R. (2005) *Repeating ourselves: American minimal music as cultural practice*. Berkeley; Los Angeles; London: University of California Press.

Ghitza, O. (2013). The theta-syllable: A unit of speech information defined by cortical function. *Frontiers in Psychology, 4.* doi:https://doi.org/10.3389/fpsyg.2013.00138

Graber, E., Simchy-Gross, R., & Margulis, E. H. (2017). Musical and linguistic listening modes in the speech-to-song illusion bias timing perception and absolute pitch memory. *Journal of the Acoustic Society of America, 142*(6), 3593–3603.

Haegens, S., & Zion Golumbic, E. (2017). Rhythmic facilitation of sensory processing: A critical review. *Neuroscience & Biobehavioral Reviews, 86,* 150–165. doi:https://doi.org/10.1016/j.neubiorev.2017.12.002

Handel, S. (1989). Listening: An introduction to the perception of auditory events. Cambridge: MIT Press.

Large, E. W., & Snyder, J. S. (2009). Pulse and meter as neural resonance. *Annals of the New York Academy of Sciences, 1169*, 46–57.

Lartillot, O., & Toiviainen, P. (2007). A MATLAB toolbox for musical feature extraction from audio. Paper presented at the 10th International Conference on Digital Audio Effects, Bordeaux. Retrieved from http://dafx.labri.fr/main/papers/p237.pdf

London, J. (2002). Cognitive constraints on metric systems: Some observations and hypotheses. *Music Perception, 19*(4), 529–550.

Longuet-Higgins, H. C., & Lee, C. S. (1982). The perception of musical rhythms. *Perception, 11*(2), 115–128.

Margulis, E. H. (2014). On repeat: How music plays the mind. New York: Oxford University Press.

Margulis, E. H., & Simchy-Gross, R. (2016). Repetition enhances the musicality of randomly generated tone sequences. *Music Perception, 33*(4), 509–514.

Margulis, E. H., Simchy-Gross, R., & Black, J. L. (2015). Pronunciation difficulty, temporal regularity, and the speech-to-song illusion.

Frontiers in Psychology, 6. doi:https://doi.org/10.3389/fpsyg.2015.00048

Mas-Herrero, E., Zatorre, R. J., Rodriguez-Fornells, A., & Marco-Pallares, J. (2014). Dissociation between musical and monetary reward responses in specific musical anhedonia. *Current Biology, 24,* 699–704.

McDermott, J. H., Schemitsch, M., & Simoncelli, E. P. (2013). Summary statistics in auditory perception. *Nature Neuroscience, 16*(4), 493–498.

Narmour, E. (1992). The analysis and cognition of basic melodic structures: The implication-realization model. Chicago: University of Chicago Press

Nozaradan, S. (2014). Exploring how musical rhythm entrains brain activity with electroencephalogram frequency-tagging. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 369*(1658), 20130393.

Ockelford, A. (2005). Repetition in music: Theoretical and metatheoretical perspectives. London: Ashgate Press.

Overath, T., McDermott, J., Zarate, J. M., & Poeppel, D. (2015). The cortical analysis of speech-specific temporal structure revealed by responses to sound quilts. *Nature Neuroscience, 18,* 903–911.

Palmer, C., & Krumhansl, C. L. (1990). Mental representations for musical meter. *Journal of Experimental Psychology: Human Perception and Performance, 16*(4), 728–741.

Patel, A. D. (2003). Language, music, syntax and the brain. *Nature Neuroscience, 6*(7), 674.

Rahn, J. (1993). Repetition. *Contemporary Music Review, 7,* 49–57.

Schaeffer, P. (1952). A la Recherche d'une Musique Concrète [In search of musique concrète]. Paris: Editions du Seuil.

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

Teng, X., Tian, X., & Poeppel, D. (2016). Testing multi-scale processing in the auditory system. *Scientific Reports, 6,* 34390.

Tierney, A., Dick, F., Deutsch, D., & Sereno, M. (2013). Speech versus song: Multiple pitch-sensitive areas revealed by a naturally occurring musical illusion. *Cerebral Cortex, 23*(2). doi:https://doi.org/10.1093/cercor/bhs003

Tierney, A., & Kraus, N. (2015). Neural entrainment to the rhythmic structure of music. *Journal of Cognitive Neuroscience, 27*(2), 400–408.

Vanden Bosch der Nederlanden, C. M., Hannon, E. E., & Snyder, J. S. (2015). Everyday musical experience is sufficient to perceive the speech-to-song illusion. *Journal of Experimental Psychology: General, 144*(2), e43.

Wen, O. X., & Krumhansl, C. L. (2017). Real time responses to Stravinsky's *Symphonies of Wind Instruments*: Perception of internal repetition and musical interest. *Music Perception, 35*(1), 60–76.

Zatorre, R. J., Belin, P., & Penhune, V. B. (2002). Structure and function of auditory cortex: music and speech. *Trends in Cognitive Sciences, 6*(1), 37–46.