Multistable perception of ambiguous melodies and the role of musical expertise

Nicholaus P. Brosowsky, and Todd A. Mondor

Citation: The Journal of the Acoustical Society of America **140**, 866 (2016); doi: 10.1121/1.4960450
View online: https://doi.org/10.1121/1.4960450
View Table of Contents: https://asa.scitation.org/toc/jas/140/2
Published by the Acoustical Society of America

Multistable perception of ambiguous melodies and the role of musical expertise

Nicholaus P. Brosowsky[a)]
Department of Psychology, The Graduate Center of the City University of New York, 365 5th Avenue, New York, New York 10016, USA

Todd A. Mondor
University of Manitoba, Winnipeg, Manitoba, R3T 2N2, Canada

(Received 14 October 2015; revised 14 July 2016; accepted 24 July 2016; published online 9 August 2016)

Whereas visual demonstrations of multistability are ubiquitous, there are few auditory examples. The purpose of the current study was to determine whether simultaneously presented melodies, such as underlie the scale illusion [Deutsch (1975). J. Acoust. Soc. Am. **57**(5), 1156–1160], can elicit multiple mutually exclusive percepts, and whether reported perceptions are mediated by musical expertise. Participants listened to target melodies and reported whether the target was embedded in subsequent test melodies. Target sequences were created such that they would only be heard if the listener interpreted the test melody according to various perceptual cues. Critically, and in contrast with previous examinations of the scale illusion, an objective measure of target detection was obtained by including target-absent test melodies. As a result, listeners could reliably identify target sequences from different perceptual organizations when presented with the same test melody on different trials. This result demonstrates an ability to alternate between mutually exclusive percepts of an unchanged stimulus. However, only perceptual organizations consistent with frequency and spatial cues were available and musical expertise did mediate target detection, limiting the organizations available to non-musicians. The current study provides the first known demonstration of auditory multistability using simultaneously presented melodies and provides a unique experimental method for measuring auditory perceptual competition.
[http://dx.doi.org/10.1121/1.4960450]

[JFL] Pages: 866–877

I. INTRODUCTION

Our sensory experience of the world is often stable and subjectively complete; we believe the world is as we perceive it. Ambiguous figures however demonstrate a disconnect between the physical world and our perceptual experience. Such stimuli are said to be multistable (Attneave, 1971) because the physical characteristics remain constant, yet observers alternate between two or more mutually exclusive perceptual interpretations (Pressnitzer and Hupé, 2006; Toppino, 2003). For example, the Necker cube (Necker, 1832) could be perceived as if the observer is above or below the cube depending on how depth cues are interpreted. Similarly, an alternating sequence of high- and low-pitched tones may be heard as two separate melodies or as a distinct single melody (referred to as the auditory streaming paradigm; van Noorden, 1975). Ambiguous figures highlight the active, inferential perceptual processes necessary for organizing noisy and often incomplete sensory information into a meaningful representation, and unsurprisingly have become an important tool for researchers studying perception. Unfortunately, although studies of the perceptual processes involved in interpreting multistable figures are ubiquitous in the visual domain (for reviews, see Leopold and Logothetis, 1999; Long and Toppino, 2004), there are very few such investigations in audition (for reviews, see Pressnitzer *et al.*, 2011; Schwartz *et al.*, 2012).

In the current study, we investigate one possible instance of auditory multistability called the "scale illusion" (Deutsch, 1974, 1975). Previous research has suggested that the scale illusion is analogous to a visual ambiguous figure (Radvansky *et al.*, 1992; Smith *et al.*, 1982), but whether the scale illusion can elicit multiple perceptual organizations has never been demonstrated. The first aim of the present work was to investigate claims made by Smith *et al.* (1982) that the scale illusion is an ambiguous figure. A secondary aim was to examine the role of intention and musical experience in supporting multistable perception. Specifically, we were interested in whether the knowledge and priming of alternative organizations will allow listeners to hear those alternatives and whether musical experience plays a mediating role.

In classic studies of the scale illusion, ascending and descending major scales are presented dichotically with successive notes from each scale alternating between ears (see Fig. 1). The stimulus sequence involves presenting the first, third, fifth, seventh, ninth, etc., notes of an ascending or descending pattern to one ear, and the second, fourth, sixth, eighth, etc., notes to the other ear. Importantly, listeners typically fail to report hearing the ascending and descending scales or patterns based on ear of input. Instead the majority of listeners report two wave-like patterns separated by

[a)]Electronic mail: nbrosowsky@gradcenter.cuny.edu

 0001-4966/2016/140(2)/866/12/$30.00

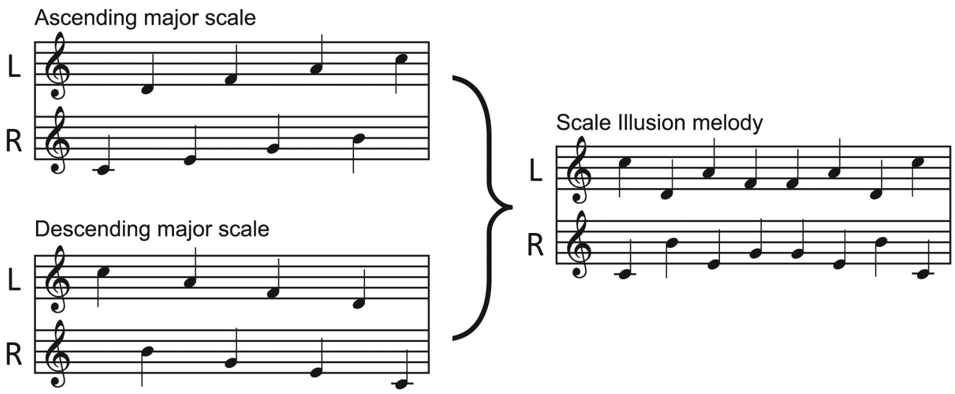

FIG. 1. An illustration of how the scale illusion melody is constructed. The scale illusion melody is the dichotic presentation of an ascending and descending major scale alternating between ears on each successive note. The ascending and descending major scales are presented in isolation on the left, the combination of which produces the scale illusion melody (on the right).

frequency range with each pattern localized to one ear [see Figs. 2(A) and 2(B); Butler, 1979; Deutsch, 1974, 1975; Judd, 1979; Radvansky *et al.*, 1992; Smith *et al.*, 1982]. In Deutsch's (1974, 1975) original study, for example, 70% of participants reported hearing both an upper and lower wave-like melody while the remaining 30% reported hearing only one wave-like melody. Additionally, right-handed listeners misperceived the locations of the melodies reporting the upper wave-like melody entirely in the right ear and the lower wave-like melody entirely in the left ear. Deutsch suggested that the complex sound sequence caused perceptual confusion and to reconcile the ambiguity the notes were organized in the simplest possible manner, which in this case is by pitch.

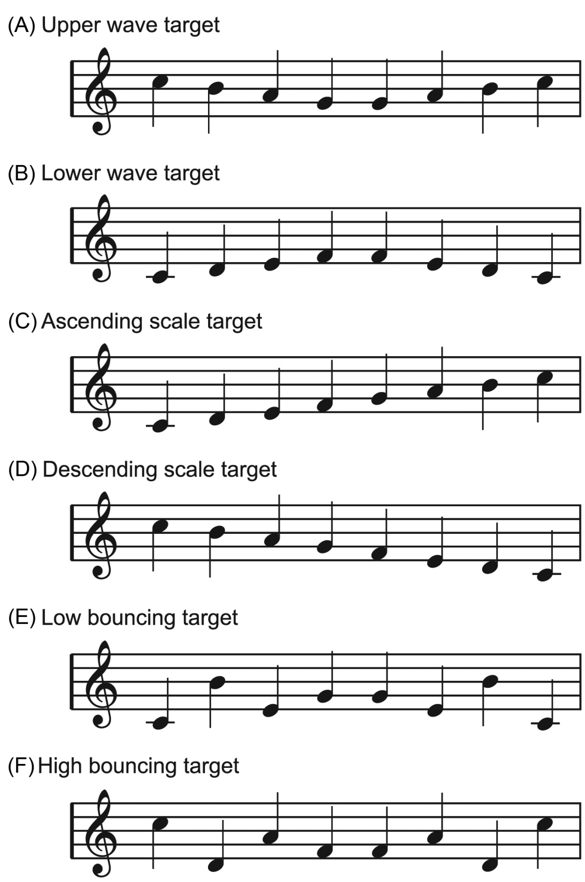

FIG. 2. Target sequences: upper (A) and lower (B) wave targets, ascending (C) and descending (D) scale targets, and low (E) and high (F) bouncing targets. All figures represent stimuli in C major.

Follow-up studies of the scale illusion largely focused on whether manipulating physical properties of the melodies presented to listeners would alter reported perceptions (Butler, 1979; Judd, 1979; Radvansky *et al.*, 1992; Smith *et al.*, 1982). The results of these studies were mixed. Some found the wave-like perceptions reported by Deutsch persisted despite manipulation of features such as intensity (e.g., Butler, 1979), while others found perceptions varied substantially depending on manipulations of spatial organization, timbre, intensity (Smith *et al.*, 1982), and note structure (Radvansky *et al.*, 1992). Still, responses did not appear to vary systematically across manipulations and there is no clear pattern of results across studies. There are, however, two common claims that emerge from this earlier work that motivates the current study. To set the stage, we briefly review each of these claims below.

The first claim is that the scale illusion is best characterized as an ambiguous figure (Smith *et al.*, 1982; Radvansky *et al.*, 1992). On this view, the spatial and pitch cues provide conflicting organizational information and because pitch happens to be the dominant cue, the spatial cues are ignored giving rise to the wave-like melodies. However, a defining characteristic of an ambiguous figure is the availability of multiple perceptual interpretations (Attneave, 1971; Toppino, 2003). All the previous studies used similar methodologies in which participants were presented with a stimulus and required to report their perception either verbally or with a forced-choice questionnaire. The use of these techniques is based on the assumption that an individual will only hear a single, clearly defined response option. This approach, of course, does not take into account the possibility that multiple perceptions are not only competing, but may all be available simultaneously to an individual in a given instance.

The work by Smith *et al.* (1982) highlights the problematic nature of this methodology. They found reported perceptions varied considerably as a function of how many response options were given and whether the participants received practice listening to the response options. In fact, the results of Deutsch (1974, 1975) could only be replicated when participants were not given practice and the forced-choice questionnaire was limited to three response options. In discussing the implications of this result they note the possibility of perceptual multistability: "It is puzzling to note the effects of practice… One explanation is that… having been alerted to other possible organizations of the sounds, subjects found their perceptions changing over time" (p. 459). Their explanation seems plausible given research on the effects of intention and prior experience on stream segregation. For example, the rate of perceptual reversals in the auditory streaming paradigm has been shown to be modulated by the intent of the listener to hear one or two melodies (Pressnitzer and Hupé, 2005, 2006; van Noorden, 1975), and prior presentation of relatively unambiguous stimuli can bias streaming (Snyder *et al.*, 2008). Dowling (1978) also explored the role of prior experience by asking listeners to identify a familiar melody interleaved with distractors. The likelihood of identifying the familiar melody increased as a function of frequency separation between the two melodies as predicted by the auditory streaming paradigm. However, when participants were told the name of the melody (e.g., "Twinkle, Twinkle, Little Star"), and even when there was no frequency separation between the melody and distractors, performance was well above chance (Dowling, 1978; Dowling *et al.*, 1987). More recently, Denham *et al.* (2014) trained participants to identify six embedded patterns in a traditional auditory streaming paradigm. After training, listeners reported spontaneous switches between all six learned alternatives. Taken together, these results suggest that knowledge of, and familiarity with, alternative organizations can modulate the segregation of auditory streams and allow listeners to hear organizations previously thought unavailable.

Similarly, the influence of prior experience on reconciling ambiguity in visual figures is well-known. For example, one line of work has demonstrated familiarity or knowledge of reversibility as critical to whether an observer will report alternative interpretations; if the observer is unaware of the reversibility of a figure, few if any alternative organizations are reported (Girgus *et al.*, 1977; Rock *et al.*, 1994). Additionally, prior experience with unambiguous versions (Botwinick, 1961), verbal information (Goolkasian and Woodberry, 2010), and contextual cues (Goolkasian, 1987) are all known to influence the initial interpretation of ambiguous figures (for a review, see Long and Toppino, 2004).

Returning to the finding of Smith *et al.*, the variability in response patterns could then be interpreted as evidence of top-down influences on perception. That is, when participants were given practice listening to all the alternatives or presented with more than three response options, alternative perceptual organizations may have become available. Such effects would have resulted in a more varied distribution of responses and an overall reduction in the apparent power of the "illusion." Despite the plausibility of such an interpretation, however, whether listeners can hear multiple organizations or alternate between them has never been tested directly.

The second claim is that musicians are more "susceptible" to the scale illusion than non-musicians reflecting a general tendency for musicians to organize sounds by pitch (e.g., Smith *et al.*, 1982). Early studies of the scale illusion only used musicians as participants (Butler, 1979; Deutsch, 1974, 1975; Judd, 1979). However, Smith *et al.* (1982) found that musicians report hearing the wave-like percepts more often than non-musicians. Furthermore, Smith *et al.* showed that whereas musicians consistently reported hearing the wave-like percepts despite variations in acoustic attributes, non-musicians were more sensitive to such variations. Similarly, other studies have found musical expertise mediates performance in auditory stream segregation tasks. For example, musicians generally report less difficulty segregating interleaved melodies (Marozeau *et al.*, 2013), can often separate streams of notes closer in pitch than non-musicians (Vliegen and Oxenham, 1999), and are better able to separate concurrent sounds (Zendel and Alain, 2009). Fundamental differences in the way musicians and non-musicians' segregate auditory streams may therefore underlie the reported differences in perceptions of the scale illusion.

Alternatively, the reported differences between musicians and non-musicians could also reflect unintentional task difficulty associated with forced-choice questionnaires and verbal reports, both of which require that participants transpose stimuli across modalities. During the forced-choice task for example, participants are typically given visual diagrams of the response options and consequently must transpose the visual stimulus (a diagram of the response option) to an auditory representation, and an auditory stimulus (the melody) to a visual representation (one associated with the response options). Musicians will almost certainly have more experience transposing stimuli between modalities and it is possible any differences reported between musicians and non-musicians could be due to this ability rather than to any fundamental differences in perceptual organization.

Therefore, whether individual listeners can hear multiple perceptual organizations of the scale illusion melody remains an open question and the role of musical expertise in producing the scale illusion remains unclear. In the present study, we investigate whether musicians and non-musicians can hear multiple, mutually exclusive organizations of dichotic melodies. To obtain an objective measure of perception, we designed a novel hidden melody recognition task similar to the interleaved melody recognition tasks used by Bey and McAdams (2002), Dowling (1978), and Dowling *et al.* (1987). Participants were presented with target sequences derived from possible perceptual organizations of the patterns typically used in studying the scale illusion (Fig. 2) and required to report the presence or absence of these targets in subsequent complex melodies, hereafter referred to as test melodies (Fig. 3). Therefore, prior to each test melody, participants had knowledge of a particular perceptual organization and were primed to listen for it. Two different test melodies in which the target sequences were technically present were tested to determine whether multistability would be unique to the scale illusion or more generally characteristic of simultaneously presented melodies overlapping in pitch and space. The first was the original scale illusion melody, here referred to as the Deutsch melody [Fig. 3(A)], and the second was a similar melody known to elicit spatially defined perceptual organizations, here referred to as the lateralized scales melody [Fig. 3(B); Radvansky *et al.*, 1992]. Importantly, test melodies in which the target sequences were *not* present were included to determine if participants could accurately discriminate between melodies that did and did not incorporate the target sequence [Figs. 3(C), 3(D), and 3(E)]. All targets

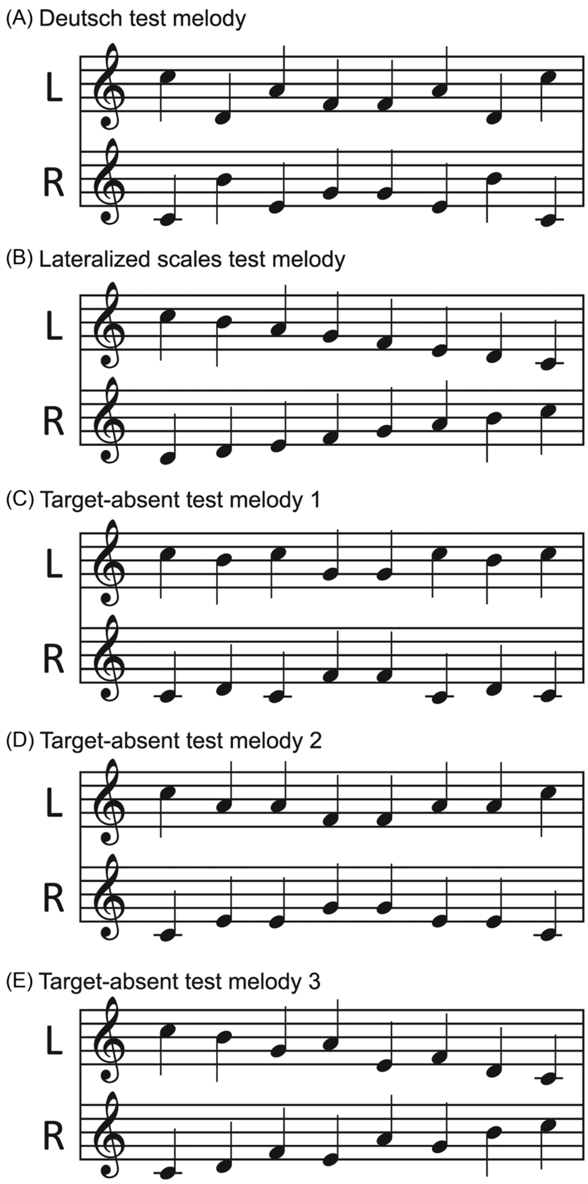

FIG. 3. Test melodies: The Deutsch test melody (A), the lateralized scale test melody (B), and the target-absent test melodies (C)–(E). All figures represent stimuli in C major.

were technically present in both the Deutsch and lateralized scales melodies. Therefore, target-absent melodies provide an important objective measure of whether participants could accurately detect the presence of the target sequences. Using a repeated-measures design, all participants had an opportunity to listen for and identify all the possible target sequences. We reasoned that if participants can reliably identify targets from mutually exclusive organizations [Figs. 2(A/B), 2(C/D), and 2(E/F)] when presented with the same melody on different occasions, then we can conclude that listeners have multiple organizations available and can alternate between them. If, on the other hand, only the most salient organization is available, then only one set of target sequences will be reliably detected for each given test melody.

It is important to note that our procedure departs from Deutsch's (1974, 1975) original procedure in a number of critical ways. The original study had participants listen to the scale illusion for a fixed length of time and subsequently report their perceptions via verbal reports or forced-choice questionnaires. In our procedure, participants are primed with a target melody and asked to report the presence or absence of that melody in a subsequent test melody. Additionally, participants in our study could listen to each test melody until they were ready to make a response (up to 20 repetitions) and made the present/absent decision while the melody was playing, not retrospectively. There are a number of benefits to the use of the hidden melody recognition paradigm that are worth highlighting. First, it allows for multiple observations from the same participants. Studies that adopted self-reports or forced-choice questionnaires typically could only collect one or two observations from each participant. Additionally, the reported perceptions of the scale illusion consisted of two complimentary melodies (the upper and lower wave-like melodies) and often subjects could only retrospectively report the presence of one or the other (e.g., Smith *et al.*, 1982). Our design allows us to assess perceptions of both the upper and lower wave-like melodies independently, providing a more accurate picture of participant perceptions. Previous work has also demonstrated that reported perceptions can vary depending on whether participants have knowledge of the alternative organizations (e.g., Smith *et al.*, 1982). In our design, participants do not have knowledge of the alternative organizations. In each block of trials, participants were only given a single target melody to search for. They were not informed that each of the target melodies was an alternative organization of the test melodies or that they would be hearing the same test melodies throughout the experiment. It would only be by the final block of the experiment that participants would have heard all the possible alternatives and we randomized the block order to control for such effects. Finally, as noted above, by including target-absent trials we can obtain an objective measure of participants' ability to hear and detect target melodies.

II. METHODS

A. Participants

A total of 52 participants (30 males and 22 females), none of whom reported any hearing impairment, took part in this study. All participants were first-year university students (approximate ages 18–22) who participated in the experiment in exchange for partial course credit in the Introduction to Psychology course at the University of Manitoba. To ensure an equal number of musicians and non-musicians, participants volunteered for separate studies under the prerequisite of either "more than four years musical training" or "less than four years musical training." In Deutsch's (1974, 1975) original study, 85% of right-handed participants reported the presence of both wave-like melodies as compared to 50% of left-handed participants. Though not all studies of the scale illusion found such effects (Judd, 1979; Smith *et al.*, 1982), to control for any possible differences in performance by right- and left-handed participants, only right-handed participants were included in the analysis.

Accordingly, four participants were excluded from the analysis.

B. Apparatus and stimuli

A computer program written using E-Prime (Psychology Software Tools, Inc., 2002) was used to present sounds and record responses. The program was run on a Dell Precision T5400 PC desktop computer (Dell, USA). Instructions, as well as any other visual information, were presented on a 22 in. LCD monitor. Sound sequences were presented over Sony MDR-600 headphones (Sony, Thailand) at a comfortable listening volume (~70 dB).

All sound sequences were created using Adobe Audition 1.5, 2004. All tones used were sinusoids of equal amplitude, and 250-ms in duration including 5-ms onset and offset amplitude ramps to eliminate onset and offset clicks. Tonal sequences had no silent gaps between consecutive tones within or between the sequences. Sound sequences were categorized as either test melodies or target sequences.

1. Test melodies

Three types of test melodies were constructed for the current study. First, the Deutsch test melody (identical to Deutsch,1975) involved ascending and descending major diatonic scales presented simultaneously beginning in opposite spatial locations and alternating location on each successive note [as shown in Fig. 3(A)]. Second, the lateralized scales test melody involved ascending and descending scales presented simultaneously, but each scale was spatially separated and presented to a single location [see Fig. 3(B)]. The lateralized scales melody has been shown to elicit percepts reflecting spatial organization rather than frequency (Radvansky *et al.*, 1992). Third, target-absent test melodies were created for each pair of target sequences in which the target sequences were *not* present. Target-absent melodies were created in the same key as the target sequence and followed a similar structural composition as the Deutsch and lateralized scale melodies in that all had simultaneous ascending and descending overlapping melodies that began on the root note of the scale and ended an octave away [see Figs. 3(C), 3(D), and 3(E)]. All test melodies were presented in both right/left and left/right spatial organizations.

2. Target sequences

The reported perceptions from previous studies included pairs of complimentary melodies (e.g., the upper and lower wave-like melodies). Therefore, three pairs of target sequences were created based on potential mutually exclusive note organizations of the Deutsch melody [Figs. 3(A/B), 3(C/D), and 3(E/F)]. Given that the lateralized scales melody contains notes identical to the Deutsch melody, target sequences also reflect potential mutually exclusive organizations of the lateralized scales melody. The first set of sequences are referred to as scale target sequences and were created so that they would only be heard if the participant had an accurate perception of the Deutsch melody using good continuity cues [Figs. 3(A) and 3(B)]. The second set of sequences were created such that they would only be heard if the participant perceived the Deutsch melody based on frequency separation [Figs. 3(C) and 3(D)]. These are the wave-like perceptions reported by Deutsch (1975) and are referred to as wave target sequences. The final set of sequences are referred to as bouncing target sequences and were created such that they would only be heard if participants perceived the Deutsch melody based on spatial location cues [Figs. 3(E) and 3(F)]. Of interest was whether participants could identify targets from more than one mutually exclusive organization when presented with the same melody. Therefore, differences between complimentary targets were not of interest, and the analyses reported below are collapsed across them.

To account for any possible carry-over effects between blocks of trials, each block was presented in a different musical key and the order of trial blocks were randomized across participants. Therefore, each pair of target sequences was created in a different key: C major (C D, E, F, G, A, B, C), G major (G, A, B, C, D, E, F#, G), and D major (D, E, F#, G, A, B, C#, D), and test melodies were created in all three keys: C major, G major, and D major. All tones were taken from an equal-tempered scale with the standard of A = 440 Hz rather than the standard used by Deutsch (1974, 1975) of A = 435 Hz. The frequencies used (in hertz) were G = 196, A = 220, B = 247, C = 262, D = 294, E = 330, F = 349, F# = 370, G = 392, A = 440, B = 494, C = 523, C# = 554, and D = 587.

3. Musical expertise questionnaire

A short questionnaire about musical training was presented to participants following the experiment to ensure that they adhered to the experimental prerequisites. Musicians were defined as those participants with at least 4 years of formal or informal musical training.

4. Procedure

Participants were required to read and sign a consent form. Following this, the researcher provided participants with an overview of the types of sound sequences they would be presented with and the types of judgments they would be required to make throughout the experiment. Next, instructions and a brief overview of the general trial structure were displayed on the computer screen. Participants were instructed to remember the target sequence presented on each trial and indicate whether it was present within a subsequent test melody. Participants were allowed to listen to the test melody as many times as they wished (up to a maximum of 20 repetitions). They were asked to answer as accurately as possible by pressing a key on a keyboard positioned directly in front of them (1 for present; 2 for absent).

Within each experimental block, participants were required to determine whether a single specific target sequence was present in each of 36 test melodies (12 trials for each of 3 test melodies). There were 8 experimental blocks for a total of 256 trials. At the beginning of each block of trials, participants were presented with the target sequence repeated 5 times (a total duration of 10 s), followed by a test melody (either the Deutsch melody, lateralized scales melody, or target-absent melody). Each trial began with the option to hear the target sequence again (another repetition of five cycles) or to skip to the test melody. Once the test melody began, participants were instructed to press 1 on the keyboard if they believed the target sequence was embedded in the test melody or to press 2 if they believed the target sequence was not embedded in the test melody. The test melody repeated up to a maximum of 20 repetitions (a total duration of 40 s) or until a response was given. The order in which trial blocks were completed was randomized across participants and trial sequences were randomized.

III. RESULTS

Two complimentary target sequences were created for each mutually exclusive perceptual organization [see Figs. 3(A/B), 3(C/D), and 3(E/F)]. Although not of theoretical importance, no significant differences were found between complimentary target sequences, $p > 0.05$ and trials were collapsed across target sequence pairs for all following analyses.

Mean percentages of "present" responses for each subject were submitted to a mixed-design, repeated measures analysis of variance (ANOVA) with one between-subjects factor, Musical Expertise (musician and non-musician), and two within-subject factors, Test Melody (lateralized scales, Deutsch, and target-absent melodies), and Target Sequence (scale, wave, and bouncing target sequences). Musicians were defined as those participants who had received more than 4 years of formal or informal musical training. Of the 48 participants, 24 were classified as non-musicians [mean number of years of formal musical training = 0.83, standard deviation (SD) = 1.17] and 24 were classified as musicians (mean number of years of formal musical training years = 7.83, SD = 2.85). A complete summary of the Musical Expertise Questionnaire is provided in Table I.

The statistical analysis revealed a significant two-way interaction between Test Melody and Musical Expertise, $F(1, 46) = 5.59$, mean squared error (MSE) = 0.04, $p = 0.005$; as compared with non-musicians, musicians made more present responses when presented with the Deutsch and lateralized scales test melodies and fewer present responses when presented with the target-absent test melodies (see Fig. 4). Therefore, musicians were overall more accurate at detecting the presence of targets than were non-musicians. The Test Melody and Target Sequence two-way interaction was also significant, $F(4, 184) = 71.26$, MSE = 0.05, $p < 0.0001$, suggesting that response patterns for target sequences differed across the three test melodies. Additionally, significant main effects were also found for Test Melody, $F(2, 92) = 82.68$, MSE = 0.04, $p < 0.0001$, and Target, $F(2, 92) = 7.26$, MSE = 0.06, $p = 0.001$.

The two-way interaction between Test Melody and Musical Expertise suggests that musicians were better overall at discriminating the presence of target sequences than non-musicians. Given the significant difference between musicians and non-musicians, performance for each group was examined separately.

A. Musicians

Mean present responses were submitted to a repeated-measures ANOVA with Test Melody (Deutsch, lateralized scales, and target-absent melodies) and Target Sequence (scale, wave, and bouncing target sequences) as factors (see Fig. 4), resulting in a significant two-way interaction between Target Sequence and Test Melody, $F(4, 92) = 52.85$, MSE = 0.04, $p < 0.0001$. Additional main effects were found for Test Melody, $F(2, 46) = 58.74$, MSE = 0.05, $p < 0.0001$, and Target, $F(2, 46) = 3.86$, MSE = 0.05, $p = 0.027$ (see Fig. 4). The interaction between Target Sequence and Test Melody suggests that the response patterns for target sequences differed across the test melodies.

To probe this interaction, further planned analyses were conducted to address two specific questions of interest. The first addressed whether or not participants could accurately detect target sequences embedded in the Deutsch and

TABLE I. Summary of results from the Musical Expertise Questionnaire.

	Formal training (years)	Informal training (years)	Performance grade level	Theory grade level
Musician ($N = 24$)				
M	7.83	9.29	2.83	1.46
SD	2.85	2.88	3.65	2.32
Median	7.00	9.00	0.00	0.00
Min	2.00	5.00	0.00	0.00
Max	15.00	15.00	9.00	9.00
Ability to read music notation	91.67%			
Ability to transcribe by ear	62.50%			
Non-Musician ($N = 24$)				
M	0.83	1.23	0.00	0.00
SD	1.17	2.29	0.00	0.00
Median	0.00	0.00	0.00	0.00
Min	0.00	0.00	0.00	0.00
Max	3.00	3.00	0.00	0.00
Ability to read music notation	12.50%			
Ability to transcribe by ear	16.67%			

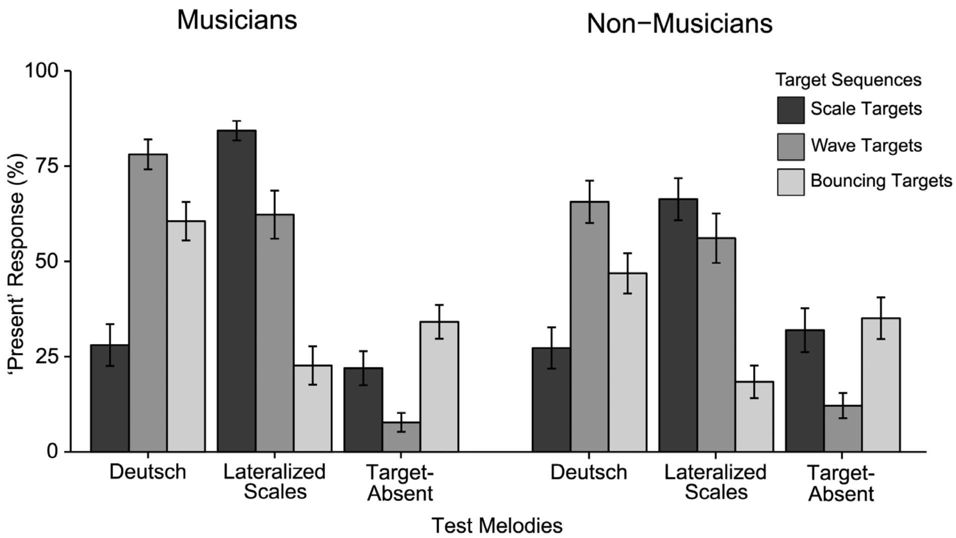

FIG. 4. Results from experiment 1. Percentage of trials where musicians (left) and non-musicians (right) reported the presence of Target Sequences (Wave, Scale, or Bouncing Targets) when listening to a Test Melody (Deutsch, Lateralized Scales, or Target-Absent).

lateralized scales melodies. The second addressed whether target salience differed in the presence of each test melody.

1. Target detection

To address the first question, planned pairwise *t*-tests were conducted comparing each target sequence, within the Deutsch and lateralized scales test melodies, against their target-absent test melody counterpart.

a. Scale target sequences. When asked to indicate whether the scale targets were present or absent, musicians responded present significantly more often in the lateralized scales test melody trials ($M = 85\%$) than the target-absent test melody trials ($M = 22\%$), $t(23) = 11.83$, $p < 0.0001$. However, there was no significant difference between responses in the Deutsch test melody trials ($M = 28\%$) and the target-absent test melody trials ($M = 22\%$), $t(23) = 1.02$, $p = 0.317$. Therefore, musicians were able to detect the presence of the scale targets when presented with the lateralized scales test melodies, but not when presented with the Deutsch test melodies.

b. Wave target sequences. When asked to indicate whether the wave targets were present or absent, musicians responded present significantly more in the Deutsch test melody trials (78%) than the target-absent test melody trials (8%), $t(23) = 12.97$, $p < 0.0001$. Similarly, musicians responded present significantly more in the lateralized scales test melodies (63%) than the target-absent test melody trials (8%), $t(23) = 7.72$, $p < 0.0001$. Therefore, musicians could accurately detect the presence of the wave targets in both the Deutsch and lateralized scales test melodies.

c. Bouncing target sequences. When asked to indicate whether the bouncing targets were present or absent, musicians responded present significantly more often in the Deutsch test melody trials ($M = 61\%$) than the target-absent test melody trials ($M = 34\%$), $t(23) = 4.67$, $p = 0.0001$. The difference between the lateralized scales test melody trials and target-absent test melody trials approached significance, however musicians made more present responses in the target-absent melody trials ($M = 34\%$) than in the lateralized scales trials ($M = 23\%$), $t(23) = -1.96$, $p = 0.06$. Musicians could therefore detect the presence of the bouncing targets when embedded within the Deutsch test melodies, but could not when embedded within the lateralized scales test melodies.

2. Target sequence salience

To address whether the salience differed between target sequences, planned pairwise *t*-tests were conducted on target sequences within the Deutsch and lateralized scales melodies.

a. Deutsch test melody. When presented within the Deutsch test melody, musicians reported the presence of the wave target sequences ($M = 72\%$) significantly more often than the bouncing target sequences ($M = 61\%$), $t(23) = 3.23$, $p < 0.001$, and significantly more often than the scale target sequences ($M = 28\%$), $t(23) = 7.48$, $p < 0.0001$. Additionally, musicians reported the presence of the bouncing target *sequence*s significantly more often than the scale target sequences, $t(23) = 3.87$, $p < 0.0001$.

b. Lateralized scales test melody. When presented with the lateralized scales test melody, musicians reported the presence of the scale target sequences ($M = 78\%$) significantly more often than the wave target sequences ($M = 63\%$), $t(23) = 3.45$, $p = 0.002$, and significantly more often than the bouncing target sequences ($M = 23\%$), $t(23) = 10.42$, $p < 0.0001$. Musicians also reported the presence of the wave target sequences ($M = 63\%$) significantly more often than the bouncing target sequences ($M = 23\%$), $t(23) = 5.71$, $p < 0.0001$.

In summary (see Table II), when presented with the Deutsch test melody, musicians could accurately report the presence of the wave and bouncing target sequences, but not the scale target sequences. Additionally, the wave target sequences were reported significantly more often than the bouncing target sequences. When presented with the lateralized scales test melody, musicians could accurately report the presence of the scale and wave target sequences, but not

TABLE II. Summary of the target detection results from experiment 1. Note: Each note organization is presented along with its organizing perceptual cues and whether the organization was available to the listener (as indicated by the target detection analysis).

		Perceptual cues			Availability	
Test melody	Target sequence	Pitch	Good continuity	Spatial	Musicians	Non-musicians
Deutsch	Wave	X	—	—	Yes	Yes
	Scale	—	X	—	No	No
	Bouncing	—	—	X	Yes	No
Lateralized scales	Wave	X	—	—	Yes	Yes
	Scale	—	X	X	Yes	Yes
	Bouncing	—	—	—	No	No

the bouncing target sequences. Furthermore, the scale targets were reported significantly more often than the wave targets. This suggests that when presented with the Deutsch and lateralized scales melodies, musicians could hear multiple mutually exclusive perceptual organizations and alternate between them.

B. Non-musicians

The analysis for the non-musician group followed the same procedure as the musician group (see Fig. 4). Mean present responses were submitted to a repeated-measures ANOVA with Test Melody (Deutsch, lateralized scales, and target-absent melodies) and Target Sequence (wave, scale, and bouncing targets) as factors resulting in a significant two-way interaction between Test Melody and Target Sequence, $F(4,92) = 24.22$, $\mathrm{MSE} = 0.05$, $p < 0.0001$, suggesting response patterns for target sequences differed across test melodies. Additionally, significant main effects for both Test Melody, $F(2,46) = 25.64$, $\mathrm{MSE} = 0.04$, $p < 0.0001$ and Target Sequence, $F(2,46) = 3.56$, $\mathrm{MSE} = 0.07$, $p = 0.04$, were found. To probe the significant two-way interaction between Test Melody and Target Sequence the same target detection and target salience analyses that were conducted for the non-musician group.

1. Target detection

a. Scale target sequences. When asked to indicate whether the scale target sequences were present or absent, non-musicians responded present significantly more often in the lateralized scales test melody trials ($M = 66\%$) than in the target-absent test melody trials ($M = 32\%$), $t(23) = 5.83$, $p < 0.0001$. However, there was no significant difference between responses in the Deutsch test melody trials ($M = 27\%$) and the target-absent test melody trials ($M = 32\%$), $t(23) = 1.29$, $p = 0.21$. Therefore, non-musicians were able to detect the presence of the scale target sequences when presented with the lateralized scales test melodies, but not when presented with the Deutsch test melodies.

b. Wave target sequences. When asked to indicate whether the wave target sequences were present or absent, non-musicians responded present significantly more often in the Deutsch test melody trials (66%) than in the target-absent test melody trials (12%), $t(23) = 7.31$, $p < 0.0001$. Similarly, non-musicians responded present significantly more often in the lateralized scales test melody trials (56%) than in the target-absent test melody trials (12%), $t(23) = 5.47$, $p < 0.0001$. Therefore, non-musicians could accurately detect the presence of the wave target sequences in both the Deutsch and lateralized scales test melodies.

c. Bouncing target sequences. When asked to indicate whether the bouncing target sequences were present or absent, non-musicians did not respond present significantly more often in the Deutsch test melody trials ($M = 47\%$) than in the target-absent test melody trials ($M = 35\%$), $t(23) = 1.63$, $p = 0.12$. The difference between the lateralized scales test melody trials and target-absent test melody trials was significant, however, indicating that non-musicians made more present responses in the target-absent test melody trials ($M = 35\%$) than in the lateralized scales test melody trials ($M = 18\%$), $t(23) = 3.29$, $p < 0.01$. Thus, non-musicians could not accurately detect the presence of the bouncing targets when presented with the Deutsch melodies or the lateralized scales test melody.

2. Target salience

a. Deutsch test melody. When presented with the Deutsch melody, non-musicians reported the presence of the wave targets ($M = 66\%$) significantly more often than the bouncing targets ($M = 47\%$), $t(23) = 2.34$, $p = 0.03$, and significantly more often than the scale targets ($M = 27\%$), $t(23) = 5.81$, $p < 0.0001$. Additionally, non-musicians reported the presence of the bouncing targets significantly more often than the scale targets, $t(23) = 3.44$, $p < 0.01$.

b. Lateralized scales test melody. When presented with the lateralized scales test melody, non-musicians reported the presence of the scale target sequences ($M = 66\%$) significantly more often than they did the bouncing target sequences ($M = 18\%$), $t(23) = 5.51$, $p < 0.0001$, though there was no significant difference between the scale ($M = 66\%$) and wave target sequences ($M = 56\%$), $t(23) = 1.05$, $p = 0.31$. Non-musicians also reported the presence of the wave target sequences ($M = 56\%$) significantly more often than the bouncing target sequences ($M = 18\%$), $t(23) = 6.49$, $p < 0.0001$.

In summary (see Table II), when presented with the Deutsch test melody, non-musicians could accurately report the presence of the wave target sequences, but not the scale or bouncing target sequences. When presented with the lateralized scales test melody, non-musicians could accurately report the presence of the scale and wave target sequences, but not the bouncing target sequences. However, there was no significant difference between the responses for the scale and wave target sequences. This suggests non-musicians had two mutually exclusive perceptual organizations available when presented with the lateralized scales test melody and could alternate between them. However, non-musicians could only accurately report a single organization when presented with the Deutsch test melody.

IV. GENERAL DISCUSSION

This study was conducted to investigate whether simultaneously presented melodies, such as are used in investigations of the scale illusion, can elicit multiple, mutually exclusive percepts and whether these perceptions are mediated by musical expertise. The results indicate that participants could identify targets from multiple, mutually exclusive organizations and alternate between them when instructed and primed to do so (for a summary, see Table II). This is the first known demonstration of multistability of simultaneously presented melodies and validates a new experimental methodology for identifying multistable melodies.

When presented with the Deutsch and the lateralized scales test melodies musicians could accurately detect two mutually exclusive organizations. Importantly, individual listeners reported the presence of targets from mutually exclusive organizations when presented with the same melodies on different occasions. This demonstrates an ability to alternate between organizations when primed to do so. In both cases, however, a third organization was unavailable to listeners. Interestingly, the available organizations aligned with either the frequency or location cues. Good continuity cues, on the other hand, were not sufficient for target identification. In the Deutsch test melody, musicians could identify the wave and bouncing target sequences, with the wave targets organized by pitch and the bouncing targets organized by spatial location. In the lateralized scales test melody, all participants could identify the scale and wave target sequences with the scale targets organized by spatial location and the wave targets by pitch. In both cases, participants were unable to identify the target sequences that did not align with at least one of these cues. Consistent with previous studies, the most salient targets when presented with the Deutsch test melody were the wave target sequences and the most salient targets when presented with the lateralized scales test melody were the scale target sequences. The most salient organizations were the same as those reported in previous studies when participants could only give one response (Deutsch, 1974, 1975; Judd, 1979; Radvansky *et al.*, 1992; Smith *et al.*, 1982) and support the interpretation that the previous results reflect a preferred organization rather than the only available organization.

With regard to the importance of musical expertise, musicians showed better detection of target sequences as they made more accurate responses when the target was present, and fewer errors when the target sequence was absent. Prior work suggested that musicians may be more likely to stream by pitch (Smith *et al.*, 1982), and our results are consistent with this in that musicians had higher response rates for the wave target sequences which are structured on the basis of pitch. However, musicians had higher response rates for all the target-present melodies while also making fewer errors on target-absent test melodies. These results suggest that musicians are not more likely to stream by pitch. Instead, musicians tend to be better overall at detecting the presence of target sequences. One possible explanation for the overall performance benefit is that musicians were better able to memorize and/or produce a mental image of previously heard melodies. Our task required participants to listen to a target sequence, remember it briefly, and compare it to a test melody. Therefore, the difference we see in performance could be the result of differences in the ability of listeners to memorize the target sequence and in auditory imagery of the target sequence during playback of the test melody. This interpretation fits with previous findings that musicians have superior auditory recognition memory (Cohen *et al.*, 2011), musicians outperform non-musicians at auditory imagery tasks (Aleman *et al.*, 2000; Janata and Paroo, 2006), and musical experience positively correlates with measures of auditory imagery (Pecenka and Keller, 2009).

It is clear however, that musicianship also had important consequences for the availability of alternative perceptual organizations. For example, musicians reliably identified the bouncing target sequence when presented with the Deutsch test melody, but non-musicians could not. Furthermore, when presented with the lateralized scales test melody, whereas musicians detected the scale target sequences significantly more often than the wave target sequences, non-musicians detected the presence of these targets equally often. It is possible in the case of the lateralized scales test melody that the experience musicians acquire as a result of repeated listening to major scales facilitated the segregation and identification of those targets. This interpretation is consistent with findings that familiarity with melodies improved identification and stream segregation in the interleaved melody tasks (Bey and McAdams, 2002; Dowling, 1978; Dowling *et al.*, 1987).

The results of the current study also have broader implications for theories of auditory stream segregation. For example, whether attention is required to segregate complex auditory stimuli into streams is still a current topic of debate (e.g., Masutomi *et al.*, 2016; Snyder *et al.*, 2012; Spielmann *et al.*, 2014; Sussman *et al.*, 2007; Thompson *et al.*, 2011). In most examinations of this issue, the standard auditory streaming paradigm is used in which alternating high- and low-pitched tones can be heard as either a single fused melody, or as two independent melodies depending on their frequency separation and the rate of the sequence. The melodies used in the current study are unique in that stream segregation is the rule rather than the exception. Specifically, there are no reports of the melodies presented in the current experiment being fused together to form a single, dyadic progression. The current study, therefore, demonstrates boundary conditions in the ability to voluntarily control perceptual grouping in cases in which stream segregation is almost a certainty. When presented with the Deutsch and lateralized scales test melodies, only organizations consistent with frequency or spatial cues were available, and prior musical experience played a mediating role in listeners' ability to discriminate the presence of those organizations.

Under one view of auditory stream segregation, top-down, schema-driven grouping is limited by bottom-up, pre-attentive processes that make use of basic auditory cues present in the stimulus (for reviews, see Bregman, 1994; Moore and Gockel, 2012). Therefore, one interpretation of the current results is that top-down processes could operate to select or reject streams, but the possible organizations available were restricted to those already established through bottom-up processing of frequency and spatial cues. Since the organizations based solely on good continuity cues were unavailable to listeners, this result would also suggest that pre-attentive processes were not able to make use of good continuity cues. There is some evidence supporting this interpretation. Bey and McAdams (2002), for example, investigated the role of schema-based segregation in streaming using an interleaved melody recognition task. Participants were presented with a target melody either before or after a melody interleaved with distractors and reported whether they were the same or different. Performance was better when the target was presented before the interleaved melody but only when there was a large mean frequency difference between the target and distractors. Bey and Adams interpreted these results as evidence that schema-driven organization could only occur after bottom-up organizations were established.

Another possibility is that attention is the limiting factor. Previous studies have demonstrated that listeners can direct their attention to specific frequency ranges (Mondor and Breau, 1999; Mondor and Bregman, 1994) and spatial locations (Mondor and Bryden, 1992; Mondor and Zatorre, 1995). Therefore, it could be the case that in order to identify the presence of the targets participants relied on the ability to search frequency ranges and spatial locations, and organizations inconsistent with at least one of these grouping cues were unavailable. On this view, the differences we see between musicians and non-musicians could reflect differences in their ability to effectively engage in attentional strategies. However, our task relied on participants to voluntarily adopt attentional strategies. Therefore, the poorer performance we see from non-musicians may not reflect a failure in applying an attentional strategy but instead a failure to determine which strategy should be adopted.

Alternatively, Jones (1976) and Jones *et al.* (1981) have suggested that stream segregation in the auditory streaming paradigm reflects an inability to direct attention with sufficient speed between successive tones with large frequency differences. It is possible that frequency and location is additive in this respect and the inability to detect those target sequences reflect an inability to switch attention with sufficient speed between tones that differ in both frequency and location. Though to our knowledge this has not been tested directly, this interpretation is consistent with a study by Mondor *et al.* (1998) who found that the selection of auditory information depended on both frequency and location similarity. Also consistent with this interpretation is our finding that musicians are capable of detecting the bouncing target sequences, arguably the most difficult in terms of frequency differences, and non-musicians could not. Therefore, the benefit musicians have in melody detection could be the result of an improved ability in switching attention between successive, expected tones.

One question for future investigation is whether multistable melodies like those used in the current study are subject to spontaneous perceptual reversals. Our findings demonstrate that listeners can identify alternatives when primed to do so, but it is unclear whether spontaneous perceptual changes occur after a particular organization is established. A number of studies, for example, have now demonstrated that the auditory streaming paradigm is subject to spontaneous reversals characteristic of visual multistability (Denham *et al.*, 2014; Pressnitzer and Hupé, 2005, 2006). Of particular interest would be whether the knowledge of alternatives and training in identifying those alternatives serve to increase the likelihood of perceptual reversals as the results of Denham *et al.* (2014) predict.

Investigations of visually ambiguous figures have found spontaneous reversals to be a robust finding with a number of defining characteristics. For example, a common finding is that the number of reported reversals increases over the viewing period, and can be reset to its original rate by simply changing the orientation of the image (e.g., Kohler, 1940). The rate of reversal is also known to be influenced by volitional strategies, previous experience with the stimulus, knowledge of alternatives, and expectations (for a review, see Long and Toppino, 2004). Ambiguous auditory figures, on the other hand, have not demonstrated the same tendency to reverse spontaneously. The verbal transformation effect has often been referred to as an ambiguous auditory figure. However, listeners typically do not spontaneously alternate between two organizations. Instead verbal transformations tend to transform into multiple alternatives that are highly idiosyncratic, and generally unpredictable (Warren, 1961; Warren and Gregory, 1958). For example, Warren (1961) presented the word "ripe" on a continuously repeating loop, which then for some transformed into "right," "rife," "ride," "life," "bright," "rape," and "wife." The Tritone Paradox involves a sequentially played pair of tones separated by a half-octave. Listeners can perceive the Tritone Paradox as either ascending or descending (Deutsch, 1987) and by manipulating the context the same listeners can alternate between perceptions. However, the Tritone Paradox does not spontaneously alternate between perceptions (Repp, 1997). Finally, listeners can intentionally alternate between interpretations of metrically ambiguous figures (i.e., repeating melodies with ambiguous beat organizations) but perceptions tend to remain stable for long periods of time before they switch, if they switch at all (Repp, 2007). From these examples, it is not clear whether spontaneous reversibility is characteristic of auditory ambiguous figures in general or unique to the auditory streaming paradigm.

Another interesting avenue for future research is the mislocalization effect. In Deutsch's (1974, 1975) original study, right-handed participants who reported the wave-like melodies also localized the upper melody entirely in the right ear and the lower melody entirely in the left ear. We did not administer an in-depth handedness questionnaire, rather relying on self-reports (Coren and Porac, 1978), and only included right-handed participants in our analysis. Therefore, we do not know if percepts varied as a function of handedness. Given Deutsch's original finding, it is possible that our results would have been different for left-handed or ambidextrous participants. Furthermore, Smith *et al.* (1982) found the mislocalization effect only for musicians; therefore we may also expect musical expertise to play a mediating role in the mislocalization effect. This issue remains unaddressed and open for future investigation.

Finally, the current study provides two novel demonstrations of auditory multistability that may be used in future studies investigating auditory stream analysis and the neural correlates of streaming. Ambiguous stimuli have become an important tool in the visual sciences for investigating perceptual organization (Leopold and Logothetis, 1999; Long and Toppino, 2004), and more recently with the advent of modern brain imaging techniques, neural correlates of conscious perception (e.g., Rees *et al.*, 2002; Sterzer *et al.*, 2009). Similarly, in the auditory domain, ambiguous figures are proving to be a useful tool in probing the mechanisms of auditory scene analysis (Bregman, 1994; Moore and Gockel, 2012; Pressnitzer and Hupé, 2005), and investigating the neural correlates of streaming (e.g., Gutschalk *et al.*, 2005; Kondo and Kashino, 2009). Recently, for example, there has been much interest in the differential activation of the "what" and "where" auditory pathways (Arnott *et al.*, 2004; Bizley and Cohen, 2013; Lomber and Malhotra, 2008). Unfortunately, although demonstrations of perceptual multistability are ubiquitous in the visual perception, there are very few such demonstrations in auditory perception (for a review, see Pressnitzer *et al.*, 2011). The stimuli created for the current study are particularly suited for dissociating such effects because the perceptual organization can be spatially or frequency directed without altering the physical properties of the stimulus. Furthermore, it seems likely that multistability is not unique to these two particular melodies but characteristic of any simultaneously presented melodies that overlap in pitch and space. Therefore, it is possible that any number of other multistable stimuli could be created using the same general framework and methodology described in the current study.

In summary, the study described above provides clear evidence that individual listeners can alternate between multiple perceptual organizations of dichotic melodies. This is the first known demonstration of auditory multistability using simultaneously presented melodies and validates a unique experimental method for measuring auditory perceptual competition. Moreover, this result confirms earlier claims (e.g., Smith *et al.*, 1982) that the scale illusion is an ambiguous figure and analogous to the visual ambiguous figures. The secondary finding was that musicians generally outperformed non-musicians at melody detection, resulting in fewer available perceptual organizations for non-musicians. This result is inconsistent with previous claims musicians are more likely to stream by pitch and instead suggests that musicians reported the wave-like melodies more often than non-musicians because of a generally superior ability to detect melodies.

Aleman, A., Nieuwenstein, M. R., Böcker, K. B., and de Haan, E. H. (**2000**). "Music training and mental imagery ability," Neuropsychologia **38**(12), 1664–1668.

Arnott, S. R., Binns, M. A., Grady, C. L., and Alain, C. (**2004**). "Assessing the auditory dual-pathway model in humans," Neuroimage **22**(1), 401–408.

Attneave, F. (**1971**). "Multistability in perception," Sci. Am. **225**, 62–71.

Bey, C., and McAdams, S. (**2002**). "Schema-based processing in auditory scene analysis," Percept. Psychophys. **64**(5), 844–854.

Bizley, J. K., and Cohen, Y. E. (**2013**). "The what, where and how of auditory-object perception," Nat. Rev. Neurosci. **14**(10), 693–707.

Botwinick, J. (**1961**). "Husband and father-in-law: A reversible figure," Am. J. Psychol. **74**(2), 312–313.

Bregman, A. S. (**1994**). *Auditory Scene Analysis: The Perceptual Organization of Sound* (MIT Press, Cambridge, MA), 773 pp.

Butler, D. (**1979**). "A further study of melodic channeling," Percept. Psychophys. **25**(4), 264–268.

Cohen, M. A., Evans, K. K., Horowitz, T. S., and Wolfe, J. M. (**2011**). "Auditory and visual memory in musicians and nonmusicians," Psychonomic Bull. Rev. **18**(3), 586–591.

Coren, S., and Porac, C. (**1978**). "The validity and reliability of self-report items for the measurement of lateral preference," Br. J. Psychol. **69**(2), 207–211.

Denham, S. L., Bõhm, T. M., Bendixen, A., Szalárdy, O., Kocsis, Z., Mill, R., and Winkler, I. (**2014**). "Stable individual characteristics in the perception of multiple embedded patterns in multistable auditory stimuli," Frontiers Neurosci. **8**, 25.

Deutsch, D. (**1974**). "An illusion with musical scales," J. Acoust. Soc. Am. **56**(S1), S25–S25.

Deutsch, D. (**1975**). "Two-channel listening to musical scales," J. Acoust. Soc. Am. **57**(5), 1156–1160.

Deutsch, D. (**1987**). "The tritone paradox: Effects of spectral variables," Percept. Psychophys. **41**(6), 563–575.

Dowling, W. J. (**1978**). "Scale and contour: Two components of a theory of memory for melodies," Psychol. Rev. **85**(4), 341–354.

Dowling, W. J., Lung, K. M.-T., and Herrbold, S. (**1987**). "Aiming attention in pitch and time in the perception of interleaved melodies," Percept. Psychophys. **41**(6), 642–656.

Girgus, J. J., Rock, I., and Egatz, R. (**1977**). "The effect of knowledge of reversibility on the reversibility of ambiguous figures," Percept. Psychophys. **22**(6), 550–556.

Goolkasian, P. (**1987**). "Ambiguous figures: Role of context and critical features," J. Gen. Psychol. **114**(3), 217–228.

Goolkasian, P., and Woodberry, C. (**2010**). "Priming effects with ambiguous figures," Attn., Percept., Psychophys. **72**(1), 168–178.

Gutschalk, A., Micheyl, C., Melcher, J. R., Rupp, A., Scherg, M., and Oxenham, A. J. (**2005**). "Neuromagnetic correlates of streaming in human auditory cortex," J. Neurosci. **25**(22), 5382–5388.

Janata, P., and Paroo, K. (**2006**). "Acuity of auditory images in pitch and time," Percept. Psychophys. **68**(5), 829–844.

Jones, M. R. (**1976**). "Time, our lost dimension: Toward a new theory of perception, attention, and memory," Psychol. Rev. **83**(5), 323–355.

Jones, M. R., Kidd, G., and Wetzel, R. (**1981**). "Evidence for rhythmic attention," J. Exp. Psychol.: Human Percept. Perform. **7**(5), 1059–1073.

Judd, T. (**1979**). "Comments on Deutsch's musical scale illusion," Attn., Percept., Psychophys. **26**(1), 85–92.

Kohler, W. (**1940**). *Dynamics in Psychology* (Liveright, New York), 158 pp.

Kondo, H. M., and Kashino, M. (**2009**). "Involvement of the thalamocortical loop in the spontaneous switching of percepts in auditory streaming," J. Neurosci. **29**(40), 12695–12701.

Leopold, D. A., and Logothetis, N. K. (**1999**). "Multistable phenomena: Changing views in perception," Trends Cognit. Sci. **3**(7), 254–264.

Lomber, S. G., and Malhotra, S. (**2008**). "Double dissociation of 'what' and 'where' processing in auditory cortex," Nature Neurosci. **11**(5), 609–616.

Long, G. M., and Toppino, T. C. (**2004**). "Enduring interest in perceptual ambiguity: Alternating views of reversible figures," Psychol. Bull. **130**(5), 748–768.

Marozeau, J., Innes-Brown, H., and Blamey, P. J. (**2013**). "The effect of timbre and loudness on melody segregation," Music Percept. **30**(3), 259–274.

Masutomi, K., Barascud, N., Kashino, M., McDermott, J. H., and Chait, M. (**2016**). "Sound segregation via embedded repetition is robust to inattention," J. Exp. Psychol.: Human Percept. Perform. **42**(3), 386–400.

Mondor, T. A., and Breau, L. M. (**1999**). "Facultative and inhibitory effects of location and frequency cues: Evidence of a modulation in perceptual sensitivity," Percept. Psychophys. **61**(3), 438–444.

Mondor, T. A., and Bregman, A. S. (**1994**). "Allocating attention to frequency regions," Percept. Psychophys. **56**(3), 268–276.

Mondor, T. A., and Bryden, M. P. (**1992**). "Orienting of auditory spatial attention: Effects of a lateralized tone cue," Neuropsychologia **30**(8), 743–752.

Mondor, T. A., and Zatorre, R. J. (**1995**). "Shifting and focusing auditory spatial attention," J. Exp. Psychol.: Human Percept. Perform. **21**(2), 387–409.

Mondor, T. A., Zatorre, R. J., and Terrio, N. A. (**1998**). "Constraints on the selection of auditory information," J. Exp. Psychol.: Human Percept. Perform. **24**(1), 66–79.

Moore, B. C. J., and Gockel, H. E. (**2012**). "Properties of auditory stream formation," Phil. Trans. Royal Soc. B **367**(1591), 919–931.

Necker, L. A. (**1832**). "LXI. Observations on some remarkable optical phenomena seen in Switzerland; and on an optical phenomenon which occurs on viewing a figure of a crystal or geometrical solid," London Edinburgh Philos. Mag. J. Sci. **1**(5), 329–337.

Pecenka, N., and Keller, P. E. (**2009**). "Auditory pitch imagery and its relationship to musical synchronization," Ann. N.Y. Acad. Sci. **1169**(1), 282–286.

Pressnitzer, D., and Hupé, J.-M. (**2005**). "Is auditory streaming a bistable percept?," in *Forum Acusticum*, Budapest, pp. 1557–1561.

Pressnitzer, D., and Hupé, J.-M. (**2006**). "Temporal dynamics of auditory and visual bistability reveal common principles of perceptual organization," Curr. Biol. **16**(13), 1351–1357.

Pressnitzer, D., Suied, C., and Shamma, S. A. (**2011**). "Auditory scene analysis: The sweet music of ambiguity," Front. Human Neurosci. **5**, 158.

Radvansky, G. A., Hartmann, W. M., and Rakerd, B. (**1992**). "Structural alterations of an ambiguous musical figure: The scale illusion revisited," Percept. Psychophys. **52**(3), 256–262.

Rees, G., Kreiman, G., and Koch, C. (**2002**). "Neural correlates of consciousness in humans," Nat. Rev. Neurosci. **3**(4), 261–270.

Repp, B. H. (**1997**). "Spectral envelope and context effects in the tritone paradox," Perception **26**, 645–666.

Repp, B. H. (**2007**). "Hearing a melody in different ways: Multistability of metrical interpretation, reflected in rate limits of sensorimotor synchronization," Cognition **102**(3), 434–454.

Rock, I., Hall, S., and Davis, J. (**1994**). "Why do ambiguous figures reverse?," Acta Psychologica **87**(1), 33–59.

Schwartz, J.-L., Grimault, N., Hupé, J.-M., Moore, B. C. J., and Pressnitzer, D. (**2012**). "Multistability in perception: Binding sensory modalities, an overview," Philos. Trans. R. Soc. B **367**(1591), 896–905.

Smith, J., Hausfeld, S., Power, R. P., and Gorta, A. (**1982**). "Ambiguous musical figures and auditory streaming," Percept. Psychophys. **32**(5), 454–464.

Snyder, J. S., Carter, O. L., Lee, S.-K., Hannon, E. E., and Alain, C. (**2008**). "Effects of context on auditory stream segregation," J. Exp. Psychol.: Human Percept. Perform. **34**(4), 1007–1016.

Snyder, J. S., Gregg, M. K., Weintraub, D. M., and Alain, C. (**2012**). "Attention, awareness, and the perception of auditory scenes," Front. Psychol. **3**, 5.

Spielmann, M. I., Schröger, E., Kotz, S. A., and Bendixen, A. (**2014**). "Attention effects on auditory scene analysis: Insights from event-related brain potentials," Psychol. Res. **78**(3), 361–378.

Sterzer, P., Kleinschmidt, A., and Rees, G. (**2009**). "The neural bases of multistable perception," Trends Cognit. Sci. **13**(7), 310–318.

Sussman, E. S., Horváth, J., Winkler, I., and Orr, M. (**2007**). "The role of attention in the formation of auditory streams," Percept. Psychophys. **69**(1), 136–152.

Thompson, S. K., Carlyon, R. P., and Cusack, R. (**2011**). "An objective measurement of the build-up of auditory streaming and of its modulation by attention," J. Exp. Psychol.: Human Percept. Perform. **37**(4), 1253–1262.

Toppino, T. C. (**2003**). "Reversible-figure perception: Mechanisms of intentional control," Percept. Psychophys. **65**(8), 1285–1295.

van Noorden, L. P. A. S. (**1975**). "Temporal coherence in the perception of tone sequences," Ph.D. thesis, Eindhoven University of Technology, Eindhoven, the Netherlands.

Vliegen, J., and Oxenham, A. J. (**1999**). "Sequential stream segregation in the absence of spectral cues," J. Acoust. Soc. Am. **105**(1), 339–346.

Warren, R. M. (**1961**). "Illusory changes of distinct speech upon repetition—the verbal transformation effect," Br. J. Psychol. **52**(3), 249–258.

Warren, R. M., and Gregory, R. L. (**1958**). "An auditory analogue of the visual reversible figure," Am. J. Psychol. **71**(3), 612–613.

Zendel, B. R., and Alain, C. (**2009**). "Concurrent sound segregation is enhanced in musicians," J. Cognit. Neurosci. **21**(8), 1488–1498.