



Cognitive Science 36 (2012) 224–260

Copyright © 2011 Cognitive Science Society, Inc. All rights reserved.

ISSN: 0364-0213 print / 1551-6709 online

DOI: 10.1111/j.1551-6709.2011.01206.x

Online Recognition of Music Is Influenced by Relative and Absolute Pitch Information

Sarah C. Creel, Melanie A. Tumlin

Department of Cognitive Science, University of California, San Diego

Received 8 February 2010; received in revised form 29 December 2010; accepted 10 March 2011

Abstract

Three experiments explored online recognition in a nonspeech domain, using a novel experimental paradigm. Adults learned to associate abstract shapes with particular melodies, and at test they identified a played melody's associated shape. To implicitly measure recognition, visual fixations to the associated shape versus a distractor shape were measured as the melody played. Degree of similarity between associated melodies was varied to assess what types of pitch information adults use in recognition. Fixation and error data suggest that adults naturally recognize music, like language, incrementally, computing matches to representations before melody offset, despite the fact that music, unlike language, provides no pressure to execute recognition rapidly. Further, adults use both absolute and relative pitch information in recognition. The implicit nature of the dependent measure should permit use with a range of populations to evaluate postulated developmental and evolutionary changes in pitch encoding.

Keywords: Music; Relative pitch; Eye tracking; Pitch perception; Pitch memory; Absolute pitch

1. Introduction

How do people recognize the sounds in their environments? One property they likely take advantage of is pitch, a fundamental dimension of sound, particularly periodic sounds such as music and speech. Like many perceptual dimensions, pitch can be encoded absolutely (its value without respect to an external standard) or relatively (its value with respect to its context). Most adult human listeners seem to process pitch primarily in relative terms: They calculate frequency ratios between successive or simultaneous pitches, rather than

Correspondence should be sent to Sarah C. Creel, Department of Cognitive Science, University of California, San Diego, La Jolla, CA 92092–0515. E-mail: creel@cogsci.ucsd.edu

processing the pitches themselves. This allows listeners to recognize both the sequence of tones 200–200–300–300 Hz and the sequence of tones 724–724–1086–1086–1086 Hz as the beginning of the “alphabet song.” Adults are not typically able to verbalize knowledge of absolute pitch (AP), but recent work by Halpern (1989), Levitin (1994), and Schellenberg and Trehub (2003; see also Trehub, Schellenberg, & Nakata, 2008) suggests that adults are actually fairly sensitive to AP in highly familiar music, although they lack pitch-labeling ability.

Another piece of information to consider in examining people’s auditory recognition is its apparent rapidity: Listeners seem to be able to recognize music on a rapid time scale. Schellenberg, Iverson, and McKinnon (1999; see also Gjerdingen & Perrott, 2008; Krumhansl, 2010) showed that listeners given a closed response set are above chance at recognizing highly familiar songs from excerpts as short as 100 ms. These studies demonstrate that some amount of information is available for recognition in very brief time intervals. This fits with a growing view of auditory processing that listeners have impressive sensitivity to detail (Palmer, Jungers, & Jusczyk, 2001; Schellenberg & Trehub, 2003), even if they cannot verbalize this knowledge.

What does it mean for music perception that listeners can recognize music with such seeming rapidity? One thing it might mean is that music recognition is *incremental*: At each instant, listeners are calculating matches between mental representations of known music and the unfolding auditory event, even before the end of the event occurs. Incremental processing accords with what we know about processing other sorts of auditory events, such as words (see Allopenna, Magnuson, & Tanenhaus, 1998; McClelland & Elman, 1986). An alternative perspective might be that listeners withhold judgment of what they are hearing until some critical amount of information has accrued, such as some sort of melodic Gestalt, but that they can guess based on incomplete information if asked to do so.

Incremental processing has been studied at length in word recognition and other domains of language processing, and it is apparent that listeners are making predictions about the likeliest continuations of what they are hearing—in a sense, they “look ahead” to upcoming material. For language, it makes sense for processing to look ahead because it should make comprehension faster, and rapid comprehension has obvious survival value. For music, the survival value is less clear, although looking ahead accords with ideas about expectancy in music (Meyer, 1967). In music, we do know that incremental processing happens over relatively coarse time scales. For instance, listeners are able to adjust their estimates of tonal stability on line (Toiviainen & Krumhansl, 2003). However, there is as yet little evidence that music recognition is incremental on brief time scales. Evidence for rapid incremental processing in a nonlinguistic domain would distinguish between rapid incremental (or predictive) processing being a general feature of cognition, versus rapid incremental processing being unique to language processing, perhaps driven by pressure to comprehend rapidly.

An interesting aspect of incremental music recognition relates to the processing of pitch information. Specifically, relative-pitch information necessarily takes slightly longer to emerge than AP information for the simple reason that relative pitch perception requires two pitches, whereas AP perception requires just one. For instance, Yankee Doodle and the

alphabet song—supposing that they were each performed in a single key all the time—could be distinguished by absolute information at the first note (e.g., C vs. A). However, they would not be distinguishable in terms of relative information until the third note, which would provide the listener with an interval of two semitones (C C D, in Yankee Doodle) or seven semitones (i.e., A A E, in the alphabet song). Thus, if using relative pitch alone to identify a melody, the listener would have to wait until a bit more information had been presented before making guesses about what they were hearing. On the other hand, in recognizing other temporally ordered events, such as words, we know that listeners begin to identify what word they think they are hearing as soon as they have any information, such as a single speech sound (e.g., Allopenna et al., 1998; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995), and do not wait for a piece of information that is not yet available (McMurray, Clayards, Tanenhaus, & Aslin, 2008). On that view, it would be surprising if listeners did not utilize information (AP) that was available sooner rather than later, presuming that they can identify that information.

1.1. *Pitch memory in human listeners*

The current picture of pitch memory in humans is a complex one. A large amount of evidence suggests that humans from infancy through adulthood are sensitive to both relative and AP information (Saffran, Reeck, Niebuhr, & Wilson, 2005) with culture-specific influences showing up as children age (Krumhansl & Keil, 1982; Lynch, Eilers, Oller, & Urbano, 1990; Trehub, Schellenberg, & Kamenetsky, 1999). For instance, 5-month-old infants are able to recognize the same melody across transpositions (i.e., changes in the AP level without changing the ratios between pitches; Chang & Trehub, 1977), and at 8 months are able to learn statistically likely sequences of pitches determined in AP alone (Saffran, 2003; Saffran & Griepentrog, 2001). Adults can recognize melodies when they are transposed (e.g., Dowling & Fujitani, 1971), but they are above chance at detecting minute (semitone) changes in AP level for familiar music (Schellenberg & Trehub, 2003). One study even suggests that listeners may improve at AP memory with age (Trehub et al., 2008).

A moment should be taken to distinguish the sort of AP perception we are considering here from the musical phenomenon of AP perception, or “perfect pitch.” In the current study, we use AP simply to refer to pitch information that is perceived or stored *without reference to a standard*. We do not use AP perception to refer to extreme *acuity* in nonrelative pitch or sensitivity to pitch chroma (recognizing E’s as different from F’s, for instance), like that in persons who possess perfect pitch (Takeuchi & Hulse, 1993), the ability to label musical notes without a reference pitch. A careful examination of demonstrations of the “implicit” (non-labeling) AP perception we are considering suggests that implicit AP does not have quite the acuity that explicit AP does (although see Ben-Haim, Chajut, & Eitan, 2010). Implicit AP is above chance for one-semitone discrepancies (58%, Schellenberg & Trehub, 2003) but improves as the discrepancy from the familiar pitch-level increases (Schellenberg & Trehub, 2003, fig. 1; see also Smith & Schmuckler, 2008; figs. 1 and 2). This is far from the accuracy level (83%) reported for explicit-AP possessors in Takeuchi and Hulse’s survey of several studies (1993, table 1), suggesting that that implicit AP has

the same mean as explicit-AP perception but a larger standard deviation—a broader tuning—than explicit-AP perception.

With respect to implicit AP knowledge, there is evidence that humans' implicit knowledge of AP may pale in comparison to other organisms, particularly avian species. For instance, Weisman, Njegovan, Williams, Cohen, and Sturdy (2004) compared three bird species (zebra finch, white-throated sparrow, and parrot), rats, and adult humans on their abilities to learn category boundaries on a pitch continuum. All three bird species dramatically outdid the two mammal species in learning categories defined by pitch range, showing sharper categorization boundaries than their human (and rat) counterparts. Work by Hulse and colleagues (Hulse, Cynx, & Humpal, 1984; MacDougall-Shackleton & Hulse, 1996) suggests that European starlings also have excellent AP perception (and can learn relative-pitch patterns). Weisman, Williams, Cohen, Njegovan, and Sturdy (2006) hypothesize that humans' lower acuity AP perception is a phylogenetic trait common across mammals. Thus, although humans show some sensitivity to AP for highly familiar stimuli (e.g., Levitin, 1994; Schellenberg & Trehub, 2003), they are evolutionarily predisposed *not* to attend to AP in recognition to the degree that bird species do.

Further, some researchers have theorized that children process pitch more absolutely but transition to more relative processing by adulthood (e.g., Saffran & Griepentrog, 2001; Sergeant & Roche, 1973; Takeuchi & Hulse, 1993). Numerous studies (e.g., Keenan et al., 2001; Sergeant, 1969) indicate that human listeners who possess pitch-labeling ability (explicit-AP perception, or "perfect pitch") almost invariably have had musical training early in life, suggesting a developmental component in pitch-labeling ability. This is consistent with an absolute-to-relative processing shift: Humans are born with sensitivity to AP but unlearn this sensitivity because it is not (usually) advantageous for recognition (Saffran & Griepentrog, 2001; Takeuchi & Hulse, 1993). What would be useful in untangling such puzzles in pitch perception is a task that could be used across a range of ages, and one that would be sensitive enough to detect varying degrees of relative and absolute encoding of pitch.

1.2. *The current study*

The present study had two major aims. The first was to ascertain whether recognition of melodies is incremental—whether listeners begin forming hypotheses about what they are hearing even before they have heard the entire event. This would bolster previous demonstrations that listeners can recognize music from very brief excerpts, and it would speak to the degree to which incrementality is unique to language processing. The second aim was to explore how listeners weight relative and AP information in identifying melodies. By establishing a benchmark with adults in a task that can potentially be used with multiple ages (and multiple species), we will not only understand adults' processing better but also provide a more stable grounding on which to explore the hypothesis that the weighting of relative and absolute information changes over time.

We employed a methodology somewhat new to music processing: eye tracking. There have been some eye-tracked explorations of music reading (e.g., Truitt, Clifton, Pollatsek,

& Rayner, 1997). However, no methodology utilizing visual fixations has been used to assess auditory music *recognition*. This method had several advantages over previous techniques. The greatest is that it can measure recognition as it is being computed—what responses listeners are considering before they make a decision, rather than just the decision itself. It is also a relatively implicit measure: Listeners make eye movements without being instructed, and no overt response is required to obtain meaningful results. The latter property also means that this methodology can be used with ages down to infancy (see McMurray & Aslin, 2004, for an example of a similar methodology with infants). Thus, if a workable eye-tracking paradigm can be developed, it can be used to examine whether there are cue-weighting shifts between AP and relative pitch over development.

Incremental recognition of words is fairly well understood. In studies of online word recognition in the “visual-world paradigm” (Cooper, 1974; Tanenhaus et al., 1995), listeners look at a picture to the extent that its label is activated by the spoken input (e.g., Allopenna et al., 1998; see McClelland & Elman, 1986, for theoretical background). For instance, listeners seeing a cat and a broom while hearing “ca . . .” would look more toward the cat and less toward the broom, even before the word ends. This is because “ca . . .” activates *cat* more strongly than it activates *broom*. On the other hand, if listeners hear “ca . . .” and see a cat and a cap, they will look equally at the cat and the cap until they have heard the end of the word (. . . t). This happens because *cat* and *cap* are equally activated by the spoken input prior to the last sound of the word.

However, listeners do not necessarily use all acoustic information that is present to distinguish words in real time. This may occur, for instance, in second-language word recognition. In one case, Weber and Cutler (2004) found that Dutch listeners recognizing English words did not use a vowel that is confusable to Dutch listeners (such as the /æ/ in “panda,” confusable with the /ɛ/ in “pencil”). Although control native-English-speaking listeners hearing “panda” used the /æ/ vowel readily, looking more to a panda than to a pencil upon reaching /æ/, Dutch listeners did not show more looks to the panda than the pencil until after the vowel, suggesting that they were not using /æ/ in recognition—they had to wait for information later in the word (. . . nda) to distinguish it from *pencil*. With respect to the current study, this means that just because a particular piece of information (such as AP) is present to distinguish stimuli, it is not necessarily the case that listeners will use that information.

We adapted this eye-tracking technique and a learning paradigm used in word recognition (Creel, Aslin, & Tanenhaus, 2008; Magnuson, Tanenhaus, Aslin, & Dahan, 2003), asking participants to learn brief melodies as “labels” for unfamiliar black-and-white shapes and then testing melody recognition by measuring looks to those shapes. Unfamiliar shapes, rather than familiar shapes or printed labels, were crucial because they deterred participants from using a secondary verbal encoding to recognize melodies. That is, listeners might associate melodies and *labels* by setting each label or shape-name to the tune of each melody (a popular technique among music history students learning to recognize classical music repertoire). If that happened, it would be difficult to determine whether any incremental processing stemmed from the music or the verbal encoding.

By having listeners learn melodic “labels” for shapes, we can assess what information factors into melody recognition without forcing listeners to use a particular type of information (APs or relative pitch). After learning, listeners saw two shapes, heard a melody associated with one of the shapes, and tried to identify the associated shape. We gauged listeners’ moment-to-moment interpretations of what melody they were hearing by measuring visual fixations to each of the associated shapes as the melody happened over time. This provided an implicit measure of listeners’ use of absolute and relative pitch during melody recognition.

Three experiments assess learners’ abilities to recognize melodies incrementally. Experiment 1 asks whether listeners use AP information in online recognition to rapidly distinguish relationally similar melodies, although relative pitch eventually distinguished them. Experiment 2 replicates Experiment 1 and addresses a potential alternate explanation—that listeners are using another relative cue, *global-relative pitch* range in the set of melodies within the experiment, to recognize melodies. Experiment 3 provides further evidence for incrementality of processing and explores the extent to which listeners use global-relative pitch information in recognition.

2. Experiment 1

How do listeners utilize AP and relative pitch information in recognition? To ask this question, we not only created a set of melodies that were distinguishable based on relative cues but also contained potentially identifying AP information. It cannot be overemphasized that *no two melodies were completely identical in relative-pitch content*. That is, listeners could perform perfectly accurately using only relative information. However, if adults spontaneously use AP information in recognition, this should be evident in their visual fixations.

To assess recognition, we used visual fixation proportions to associated shapes (examples in Fig. 1) to track the activation level of each melody’s mental representation in real time. Target pictures occurred on each trial with either the picture corresponding to the melody with the first three intervals matched (e.g., melodies a1 and a2 in Table 1), or the picture for a dissimilar melody (e.g., melodies a1 and e1 in Table 1). The two-alternative design was preferable to multi-alternative designs often used in language experiments, because pilot testing indicated that the melody-learning task became extremely difficult for listeners when there were four, rather than two, alternatives (see Swingle, Pinto, & Fernald, 1999; Spivey, Grosjean, & Knoblich, 2005, for similar two-alternative paradigms in the developmental word-recognition and mouse-tracking literatures, respectively).

If melody recognition is incremental, listeners should rapidly visually fixate the shape for a melody that is different in relative information from the other shape’s melody—they will distinguish dissimilar melodies based on relative-pitch differences, as well as rhythmic differences present between dissimilar melodies. This is like distinguishing between the alphabet song and “Happy Birthday,” which differ in relative pitch information by the third note. A second question is whether listeners also use AP information to identify melodies.

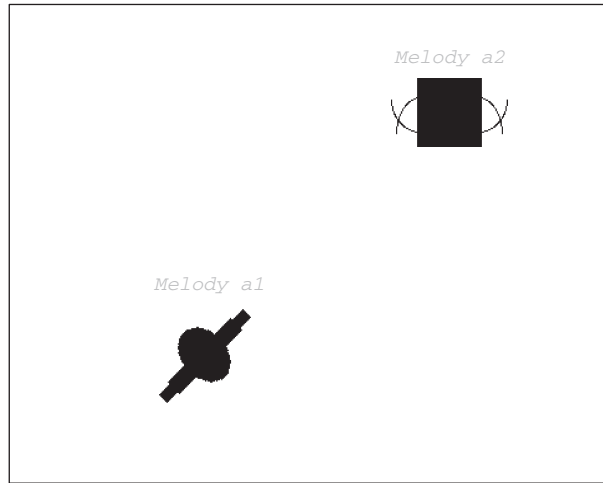


Fig. 1. Experiment 1, example trial. Text in gray specifies the melody associated with each shape and was not actually presented to participants. Participants in other conditions learned to associate different shapes with those melodies.

Use of AP information will simply lead to more rapid recognition. Note that listeners are not in any sense “forced” by our task to use AP information, because they can reach ceiling accuracy without it. This is important because we want to know whether AP information is *routinely* represented in memory and used in recognition. If they use AP information, then they should look at the correct shape sooner when two shapes’ melodies differ only in AP early on (a1 vs. a2 in Table 1 List 1, which differ at their first notes) instead of waiting until relative information (the last pitch interval) distinguishes the melodies.

Note that this methodology uses looks to associated pictures as a proxy for melody recognition. The success of this methodology is predicated on the assumption that listeners have made strong melody-shape associations. To ensure strong melody-shape associations, we trained listeners to high accuracy levels. We also used several different assignments of melodies to shapes, to guard against the possibility that certain melody-shape associations might be easier to learn than others.

In selecting AP differences for our stimuli, we considered the rather broad tuning evident in ordinary listeners’ AP knowledge (e.g., Schellenberg & Trehub, 2003; Smith & Schmuckler, 2008), as well as experimental precedent (Saffran & Griepentrog’s (2001) work on absolute vs. relational pitch processing in infants), and selected six semitones as an appropriately registerable difference. This six-semitone difference represents the approximate difference in ranges between a soprano and an alto, the two basic female voice ranges in Western music. Thus, any evidence we find of AP processing can be taken to denote somewhat coarse-grained apprehension of AP. One might reasonably presume that smaller pitch differences would create qualitatively similar, but quantitatively smaller, effects. Moreover, this six-semitone pitch difference results in two keys which are harmonically distant from one another, reducing the chances that participants would be confused in their

Table 1
Stimuli and conditions used in Experiment 1













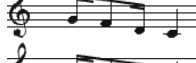
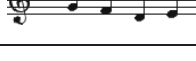

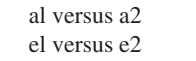
Identifier	Melody	Transposed to			
		List 1	List 2	List 3	List 4
a1		F#4^a	C4	C4	C4
a2		C4^a	F#4	C4	C4
b1		F#5	C5	C5	C5
b2		C5	F#5	C5	C5
c1		C5	F#5	F#5	F#5
c2		F#5	C5	F#5	F#5
d1		C4	F#4	F#4	F#4
d2		F#4	C4	F#4	F#4
e1		C4	C4	F#4	C4
e2		C4	C4	C4	F#4
f1		C5	C5	F#5	C5
f2		C5	C5	C5	F#5
g1		F#5	F#5	C5	F#5
g2		F#5	F#5	F#5	C5
h1		F#4	F#4	C4	F#4
h2		F#4	F#4	F#4	C4
Test Items ^b					
Pairing Type	Pictures Presented	Absolute Pitch Match			
Paired	a1 versus a2	Diff. pitch	Diff. pitch	Same pitch	Same pitch
Paired	e1 versus e2	Same pitch	Same pitch	Diff. pitch	Diff. pitch
Dissimilar	a1 versus e1	Diff. pitch	Same pitch	Diff. pitch	Same pitch
Dissimilar	a2 versus e2	Same pitch	Diff. pitch	Same pitch	Diff. pitch
Paired	b1 versus b2	Diff. pitch	Diff. pitch	Same pitch	Same pitch
Paired	f1 versus f2	Same pitch	Same pitch	Diff. pitch	Diff. pitch

Table 1
(Continued)

Test Items ^b					
Pairing Type	Pictures Presented		Absolute Pitch Match		
Dissimilar	b1 versus f1	Diff. pitch	Same pitch	Diff. pitch	Same pitch
Dissimilar	b2 versus f2	Same pitch	Diff. pitch	Same pitch	Diff. pitch

Notes. Bolded melodies and pitch levels are referred to in the text.

^aFor reference, C4 = 261.6 Hz, F#4 = 370 Hz, C5 = C4*2 = 523.3 Hz, F#5 = F#4*2 = 740 Hz.

^bEach picture only appeared with two other pictures: the one associated with the picture’s paired melody, and a specific item from another pair. The *an* and *en* shapes (a1 with e1, a2 with e2) always appeared together, as did *bn* *fn*, *cn* *gn*, and *dn* *hn*.

identification by carrying over relative-pitch information from a previous melody (Bartlett & Dowling, 1980).

2.1. Method

2.1.1. Participants

Sixteen participants from the University of California, San Diego (La Jolla, CA) reporting no AP perception abilities took part for course credit. Three additional participants were replaced: two failed to reach 90% correct in the 2-h experiment time frame, and one had too many missing data resulting from a poor eye track. For the final sample, experience playing music ranged from 0 to 16 years ($M = 7.1$, $SD = 5.1$) and did not affect eye-tracking results. Fifty percent had had one or more music theory courses, one person had had a music perception course, 31% had had music history courses, and four had had no music coursework aside from performing musical groups.

2.1.2. Stimuli

2.1.2.1. *Shapes*: The shapes which were associated with melodies have been used before in word-learning tasks (e.g., Creel, Aslin, & Tanenhaus, 2006; Creel, Tanenhaus, & Aslin, 2006; Creel et al., 2008).

2.1.2.2. *Melodies*: Melodies were created in BARFLY 1.73 software (Taylor, 1997; <http://www.barfly.dial.pipex.com/>) using the recorder timbre and a tempo of 120 beats per minute. After creation, melodies were exported as .aiff files, which were then converted to .wav files, and scaled to 70 dB for uniform loudness in PRAAT 5.1.20 software (Boersma & Weenink, 2009).

All four- to five-note melodies were constructed in pairs using pitches either from C major or F# major, in a roughly seven-semitone range above middle C (C4), F#4 (the F# above middle C), C5 (an octave above middle C), or F#5. Melodies are displayed in musical notation in Table 1 as though each were played from notes surrounding C4 (middle C). In different conditions (Lists 1–4), the actual pitch region of each melody was assigned as specified in Table 1. Across participants, each melody occurred in each possible condition (described under Procedure). Each melody spanned roughly one perfect fifth (seven semitones).

2.1.2.3. Procedure: Participants were tested individually in a sound-treated room on a Mac Mini running Matlab experimental presentation software written with Psychtoolbox (Brainard, 1997; Pelli, 1997) and the embedded Eyelink Toolbox (Cornelissen, Peters, & Palmer, 2002). Eyes were tracked at 4-ms resolution with an Eyelink Remote eye tracker (SR Research, Mississauga, ON), positioned in front of the Mac monitor. Sounds were presented via Sennheiser HD280 Pro headphones (Sennheiser Electronic Corporation, Old Lyme, Connecticut, US) adjusted to a comfortable listening level.

Participants saw the following printed instructions:

In this experiment, you'll be hearing a lot of short melodies. You'll also see lots of unfamiliar objects. Each melody goes with one particular object, kind of like a musical "word" for that object.

For each melody you hear, you'll see several pictures. Make a guess as to which picture it's a "word" for. If you're right, that picture will stay on the screen. If you're wrong, it will disappear and only the correct picture will remain onscreen.

At first you'll have to guess, but you'll slowly get better at it.

Click the mouse to continue.

Participants then took part in training and testing trials. Half of the learning and test trials were paired-melody trials. On the other half of trials, each target shape was assigned a particular shape from another pair which was associated with a *dissimilar* melody. That is, a1 appeared with a2 on paired-melody trials, and with e1 on dissimilar-melody trials (see Table 1, bottom, for more examples). Dissimilar melodies (such as a1 and e1) were designed to be as discriminable as possible while adhering to normal Western musical conventions: They diverged from each other in relative pitch, rhythm, or both within the first 250 ms of each melody. Thus, dissimilar-melody trials should be much more easily discriminated, serving as a baseline for recognition speed in the easiest case, and allowing assessment of recognition based on similarity in pitch range without similarity in relative-pitch or rhythmic information. Each picture occurred with only two other pictures, and equally often, so that frequency of co-occurrence—which human perceivers are highly sensitive to (e.g., Fiser & Aslin, 2005; Saffran, Aslin, & Newport, 1996)—was matched for similar-melody pictures (e.g., a1 and a2) and dissimilar-melody pictures (a1, e1).

Orthogonal to the paired/dissimilar factor was an AP-match factor. Half each of the paired and dissimilar trials were drawn from the same pitch range (e.g., that around middle C [C4]; in List 1, e1 vs. e2, and a2 vs. e2), and the other half were drawn from different pitch ranges (always six semitones apart; in List 1, a1 vs. a2, and a1 vs. e1). Each individual shape, for a given participant, only occurred with two other shapes: the paired-melody shape and a dissimilar-melody shape. In addition to four different pitch-level assignments (Lists 1–4), there were four different assignments of shapes to melodies, which were combined with each possible pitch-level list to yield 16 conditions.

Each block of training was 128 trials long. Trials within a block were presented randomly. On each training trial, two shapes appeared; after 500 ms, a melody played, and the listener guessed which shape went with that melody. As feedback, 200 ms after the click, only the correct shape stayed on screen after a response. The intertrial interval was 500 ms. Training continued until the participant reached 90% accuracy for a 128-trial block. There was then a screen of test instructions telling listeners that they would no longer receive feedback. The test phase (two 128-trial blocks with no pause between) was identical to the training trial blocks, except that no feedback was provided. At no point during training or test were participants asked to respond at a particular rate of speed, as this might have pressured them into processing incrementally rather than doing so naturally. Visual fixations were monitored throughout but were analyzed only for test trials.

2.2. Results

Listeners reached high levels of accuracy, recognized melodies rapidly, and appeared to use both relative and AP information to do so. Because our items (melodies) were completely counterbalanced across participants, we follow Raaijmakers' recommendation (Raaijmakers, 2003; Raaijmakers, Schrijnemakers, & Gremmen, 1999) of relying on analyses across participants to determine statistical significance. For full information we also report items analyses. Effect size is reported as generalized eta-squared (Bakeman, 2005; Olejnik & Algina, 2003) which equates effect size across within- and between-participants designs.

2.2.1. Accuracy

Participants reached accuracy criterion on all trial types in 3.56 trial blocks on average ($SD = 1.09$). Accuracy (Fig. 2) was slightly higher for dissimilar-melody trials (96%) than for paired-melody trials (94%), but different-AP trials and same-AP trials did not differ in accuracy. An analysis of variance (ANOVA) on percent correct with AP Match (same, different) and Pair Type (paired, dissimilar) as factors confirmed this. The effect of Pair Type approached significance, $F_1(1, 15) = 4.00, p = .06$; $F_2(1, 15) < 1$; $\eta^2_G = .055$, but AP Match and the AP Match \times Pair Type interaction did not (all $F_s < 1$).

2.2.2. Eye-tracking data

Overall, listeners showed evidence of recognition prior to the end of the melody. Further, listeners seemed to be affected by AP information in distinguishing melodies. Because it takes about 200 ms to plan and execute an eye movement based on external information (Hallitt, 1986), we analyzed time windows that were shifted 200 ms later than particular time points in the melody. The unique interval of a paired melody always happened at 500 ms after melody onset ($+200 = 700$), so we analyzed two time windows before this point (200–450 and 450–700) and one after (700–950). Looks executed before 200 ms were removed from analysis. Fixations were preprocessed into 50-ms bins for display and analysis.

In this and following experiments, we analyze time windows where at least 95% of trials were still ongoing, as eye tracking stopped when a response was made. To include later time windows would (sometimes dramatically) overrepresent trials where listeners

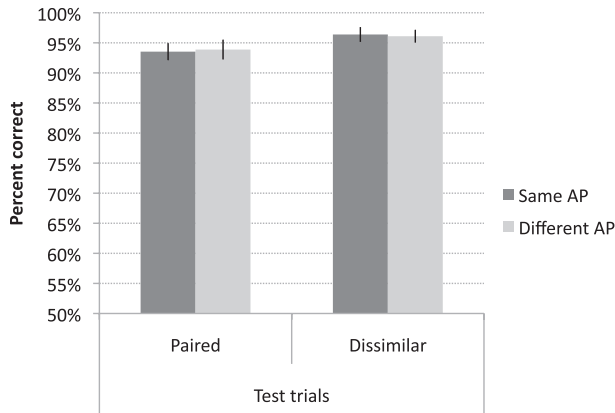


Fig. 2. Experiment 1, accuracy on test trials. Error bars are standard errors.

were, for example, less certain of their response. To see how other authors have dealt with unequal response time issue, see Allopenna et al. (1998) and Mirman and Magnuson (2009).

Figure 3 displays averaged looks toward the correct shape minus those toward the incorrect shape—the “target advantage.” When target advantage is at zero, listeners are fixating both shapes equally often at that time point, suggesting they have yet to distinguish which is correct. When target advantage exceeds zero, participants are fixating the correct shape more than the incorrect one. Target advantage increased over time from the start of the melody, with looks rising above zero beginning around 500 ms. Listeners looked faster to correct shapes when relative information distinguished melodies early (gray lines) or when AP information (solid black line) distinguished them, than when only a final pitch interval distinguished melodies (dashed black line). All but the same-AP paired melodies were recognized before the final interval.

A within-participants ANOVA on target advantage with AP Match (same, different), Pair Type (paired, dissimilar), and Time Window (200–450, 450–700, 700–950) as factors confirmed these observations. Correct looks increased over time, effect of Time Window: $F_1(2, 30) = 96.31$, $p < .0001$; $F_2(2, 30) = 54.75$, $p < .0001$; $\eta^2_G = .54$. Dissimilar-melody trials (gray lines, Fig. 3) showed more fixations to the correct shape than paired-melody trials, effect of Pair Type: $F_1(1, 15) = 4.94$, $p = .04$; $F_2(1, 15) = 1.1$, $p = .31$; $\eta^2_G = .034$. Different-AP melodies—both paired and dissimilar—were recognized more strongly than same-AP melodies in later time windows, AP Match \times Time Window interaction: $F_1(2, 30) = 5.48$, $p = .009$; $F_2(2, 30) = 3.38$, $p < .05$; $\eta^2_G = .031$. No other effects reached significance.

Because we wanted to assess use of AP information to tell apart relationally similar melodies, we conducted an ANOVA that was limited to paired trials, with AP Match and Time Window as factors. Correct looks increased over time, effect of Time Window: $F_1(2, 30) = 50.13$, $p < .0001$; $F_2(2, 30) = 36.97$, $p < .0001$; $\eta^2_G = .52$. Looks on different-AP and same-AP trials diverged over time, with an advantage for different-AP trials in later time windows, AP Match \times Time Window interaction: $F_1(2, 30) = 5.38$, $p = .01$;

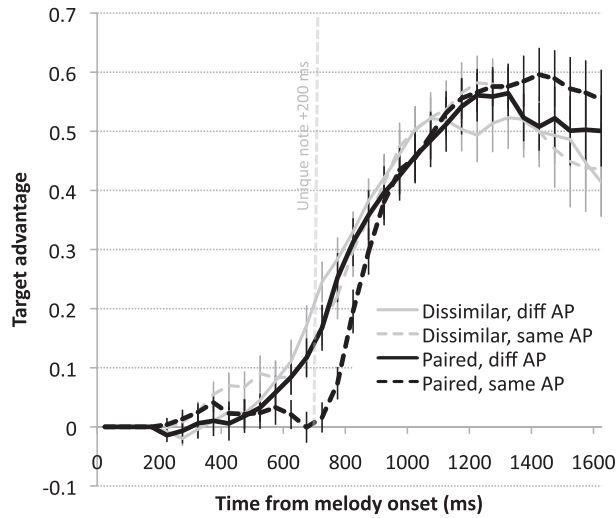


Fig. 3. Experiment 1, target advantage (correct looks minus incorrect looks) on paired-melody trials (black) and dissimilar-melody trials (gray).

$F_2(2, 30) = 4.95, p = .01; \eta^2_G = .06$. Same- and different-AP trials did not differ in the first two windows ($p \geq .14$) but did differ in the third time window, $t_1(15) = 2.77, p = .01$; $t_2(15) = 2.42, p = .03$. At a finer grain, in the second window, different-AP trials exceeded zero, $t_1(15) = 2.5, p = .02$; $t_2(15) = 1.64, p = .12$, meaning different-AP melodies were identified before the distinguishing note of the melodies could be processed. Same-AP items did not exceed zero in this time window, $t_1(15) = 0.82, p = .42$; $t_2(15) = 0.53, p = .60$, meaning they were not identified until after the unique interval had been processed. This fits with an account where listeners use AP information to recognize melodies rapidly, although relative information would soon distinguish them.

2.3. Discussion

Listeners seem to recognize melodies in an incremental fashion, much like words, looking to the correct associated shape upon hearing only part of the melody. Further, listeners looked to the associated shape more quickly when that shape and the other shape were associated with highly similar melodies differing in AP level, than when highly similar melodies had the same pitch level. This suggests that listeners not only encode AP information about music they learn in a brief experiment but also use AP in recognition even when later relative information (the final interval of each melody) is sufficient to distinguish them. AP is at least an implicit factor shaping the process of recognizing music.

However, there is an alternative, relative-pitch explanation for these results: Listeners may have been encoding not actual pitch levels of melodies, but pitch levels *relative to* the entire pitch range heard during the experiment. For instance, they might encode a melody near C4 not as C4 but as “low,” and a melody near F#5 as “high.” This is similar to

Navon's (1977) global–local distinction for visual stimuli, such as a large triangle composed of three small squares: The local information is the squares, whereas the global information is the triangle (for an auditory analog of these effects, see Justus & List, 2005). In fact, some previous research on pitch-learning effects in infants and adults (Saffran, 2003; Saffran & Griepentrog, 2001; Saffran et al., 2005) does not distinguish between true AP representation and global encoding of relative pitch range, making this a novel exploration of the issue. We will refer to this type of relative encoding as *global-relative* encoding. This is used to distinguish it from *local-relative* coding—the individual pitch intervals or contour changes within a single melody.

In Experiment 2, we dissociated global-relative encoding effects from true AP encoding effects. After training listeners on melodies, we changed AP information without altering the global-relative information. Such a change should not disrupt processing if listeners use only global-relative information. However, if listeners obligatorily encode AP information, then shifting AP should strongly affect recognition.

3. Experiment 2

3.1. Method

3.1.1. Participants

New participants ($N = 36$) from the same pool as Experiment 1 took part. Ten more were replaced: Two did not reach criterion performance, and, due to ongoing lab training on the eye tracker, 8 had significant eye-tracking data loss. Five participants did not report music data. Experience playing music ranged from 0 to 20 years ($M = 4.35$, $SD = 4.89$). Fourteen percent had had one or more music history course, one person had had a computer music course, and 69% had had no music coursework at all aside from performance groups.

3.1.2. Stimuli

3.1.2.1. Melodies: Melodies (Table 2) were created in FINALE 2009 software (2008, Make-Music, Inc., Eden Prairie, MN, USA), using a whistle timbre. They were exported from Finale as .aiff files, at a tempo of 90 beats per minute and a MIDI velocity (loudness) of 101, and were then converted to .wav files in PRAAT.

Instead of pairs of melodies, this experiment employed four *triples* of melodies (e.g., a1, a2, a3 in Figs. 4a and 4b; Table 2). In each triple, all three melodies matched in relative pitch until the last note, and two matched in AP until the last note. The third melody was six semitones lower than the other two. Which melody in the triple was the lower pitched one varied between the three pitch-assignment lists used (Table 2) but was constant for a given participant.

This odd-one-out structure created six “paired-melody” shape combinations within a triple, of which 1/3 were same-AP (e.g., a2–a3 in List 1) and 2/3 were different-AP (e.g.,

Table 2
Melodies used in Experiment 2

Melody		Transposed to					
		List 1		List 2		List 3	
		Pre	Shifted	Pre	Shifted	Pre	Shifted
a1		C4	F#4	F#4	C5	F#4	C5
a2		F#4	C5	C4	F#4	F#4	C5
a3		F#4	C5	F#4	C5	C4	F#4
b1		F#4	C5	C5	F#5	C5	F#5
b2		C5	F#5	F#4	C5	C5	F#5
b3		C5	F#5	C5	F#5	F#4	C5
c1		C4	F#4	F#4	C5	F#4	C5
c2		F#4	C5	C4	F#4	F#4	C5
c3		F#4	C5	F#4	C5	C4	F#4
d1		F#4	C5	C5	F#5	C5	F#5
d2		C5	F#5	F#4	C5	C5	F#5
d3		C5	F#5	C5	F#5	F#4	C5

Pairing Type	Pictures Presented ^a		Absolute Pitch Match				
Paired	a1 versus a2	Diff. ^b	Interf. ^b	Diff.	Diff.	Same ^b	Same
Paired	a2 versus a1	Diff.	Diff.	Diff.	Interf.	Same	Same
Paired	a1 versus a3	Diff.	Interf.	Same	Same	Diff.	Diff.
Paired	a3 versus a1	Diff.	Diff.	Same	Same	Diff.	Interf.
Paired	a2 versus a3	Same	Same	Diff.	Interf.	Diff.	Diff.
Paired	a3 versus a2	Same	Same	Diff.	Diff.	Diff.	Interf.
Dissimilar	a1 versus c2	Diff.	Interf.	Diff.	Diff.	Same	Same
Dissimilar	a2 versus c1	Diff.	Diff.	Diff.	Interf.	Same	Same
Dissimilar	a1 versus c3	Diff.	Interf.	Same	Same	Diff.	Diff.
Dissimilar	a3 versus c1	Diff.	Diff.	Same	Same	Diff.	Interf.
Dissimilar	a2 versus c3	Same	Same	Diff.	Interf.	Diff.	Diff.
Dissimilar	a3 versus c2	Same	Same	Diff.	Diff.	Diff.	Interf.

Notes. Diff. = different pitches; Interf. = interference from absolute-pitch memory.
Bolded pitch levels for a1, a2, and a3 illustrate the potential for pitch interference after the shift.
^aPictures for set a melodies only appeared with those from c, and b appeared with d.

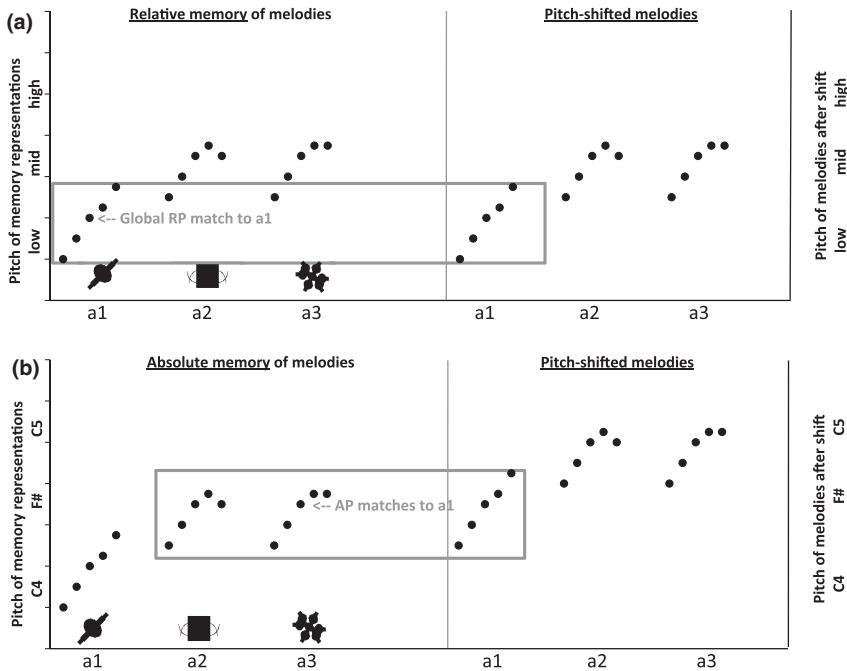


Fig. 4. Experiment 2, depiction of relative (a) or absolute (b) memory representations of melodies. Gray boxes show perceived pitch similarity if listeners have encoded only relative information (in (a)) or absolute information (in (b)). Pitch (y-axes) is in semitones, a log scale.

a3–a1, a1–a2 in List 1). As before, there were dissimilar-melody trials where melodies had different relative information and either same (e.g., a1–c1) or different (a1–c2) pitch ranges.

As before, listeners were trained to recognize melodies at 90% accuracy. Then they were tested on nonreinforced trials. The first block of test trials was at the same AP level as during the reinforced training trials. In the second block of test trials, which happened after a break in the experiment, all melodies were shifted up by six semitones (right sides of Figs. 4a and 4b; note that 4a and 4b differ crucially in their y-axis labelings). From a global-relative perspective, this should not change things at all: The highest pitch range is still the highest pitch range, and the lowest is still the lowest (Fig. 4a).

However, from an AP perspective, this should be a very confusing thing to do. Specifically, listeners should be misled on trials where the target melody is a1 (Fig. 4b). This is because the shifted version of a1 is at the AP level where a2 and a3 used to be. If listeners encoded a2 and a3 in terms of AP, then a shifted a1 should, prior to the final interval, activate the representations of a2 and a3 more strongly than a1. These trials will be referred to as *interference trials*. For other types of trials, this interference should not show up. For instance, for a2 versus a3 trials (same-AP trials), the shifted melody (either a2 or a3) is equally distant in pitch from both a2 and a3 representations. Also, for trials where a1 is the *distractor* and a2 or a3 is the target (different-AP trials), there should be little interference because shifted a2 (or a3) is closer in pitch level to the a2 and a3 representations (six

semitones below) than to a1 representations (one octave below). Note that, prior to the pitch shift, interference trials are essentially identical to the different-AP trials; it is only by shifting the pitches so that old and new AP content collide that interference might be generated.

Melodies spanned three pitch ranges (around C4, F#4, C5) before the pitch shift; after the shift, they spanned three different pitch ranges (F#4, C5, F#5). As before, there were equal numbers of paired and dissimilar trials, and the three pitch-shift lists were crossed with six melody-shape assignments to yield 18 conditions.

3.1.2.2. Shapes: These were a subset of those used in Experiment 1.

3.1.2.3. Procedure: This was similar to Experiment 1, with blocks of randomly ordered two-alternative trials (96 per block). We modified displays so that, instead of appearing in four possible locations, pictures appeared in two screen locations to the left and right of center. This was merely to simplify the counterbalancing of target location, melody, and trial type. We trained participants to 90% correct. A screen of test instructions was presented (2 s minimum), telling listeners that they would no longer receive feedback. Test blocks followed. The first was at the training pitch level, similar to Experiment 1. This verified that listeners had learned the melodies well enough to identify them, and it provided visual fixation information when trials were not being reinforced. The second test block, which was presented after a short break during which the participant was instructed to converse with the experimenter, was shifted up in pitch by six semitones. The break between blocks was also important because it reduced the possibility that working memory of previous melodies would interfere with (or aid) listeners' recognition.

3.2. Results

3.2.1. Accuracy

Listeners reached 90% or better accuracy in 4.67 blocks ($SD = 2.92$). Overall, paired-melody trials (Fig. 5, left) were less accurate than dissimilar-melody trials (Fig. 5, right), and same-AP trials (dotted lines) were less accurate than different-AP trials (solid lines). Accuracy during the test stayed relatively high from before to after the global pitch shift, though there looked to be a drop in accuracy on trials with interference (gray lines).

An ANOVA on proportion correct with AP Match, Pairing Type, and Shift (preshift, postshift) confirmed these findings. An effect of Shift, $F_1(1, 35) = 4.40$, $p < .05$; $F_2(1, 11) = 4.52$, $p = .06$; $\eta^2_G = .006$, suggested that listeners were more accurate before than after the pitch shift. An effect of Pairing Type, $F_1(1, 35) = 14.85$, $p = .0005$; $F_2(1, 11) = 13.03$, $p = .004$; $\eta^2_G = .048$, confirmed that paired-melody trials were less accurate than dissimilar-melody trials, and an effect of AP Match, $F_1(2, 70) = 19.68$, $p < .0001$; $F_2(2, 22) = 12.61$, $p = .0002$; $\eta^2_G = .10$, confirmed that same-AP trials were less accurate than different-AP trials. There was also a Shift \times AP Match interaction, $F_1(2, 70) = 4.01$, $p = .02$; $F_2(2, 22) = 3.41$, $p = .05$; $\eta^2_G = .008$. Therefore, we looked at AP Match effects for preshift and postshift trials. For both, there were significant effects of AP Match, but in

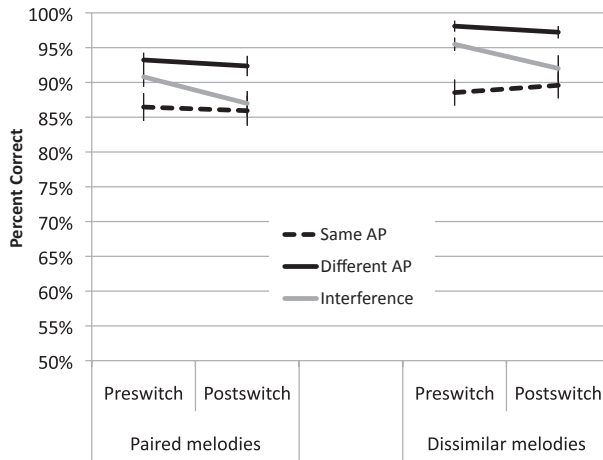


Fig. 5. Experiment 2, accuracy on test trials. Error bars are standard errors.

preshift test trials, the interference trials were more accurate than the same-AP trials, $t_1(35) = 3.37$, $p = .002$; $t_2(11) = 2.4$, $p = .04$, whereas in postshift test trials, interference trials were not more accurate than same-AP trials, $t_1(35) = 1.24$, $p = .23$; $t_2(11) = 1.24$, $p = .24$. This suggests that participants were relatively less accurate on the interference trials after the pitch shift, consistent with confusion based on pitch dissimilarity. Overall, though, accuracy remained high, suggesting that relative information was used for overt identification.

3.2.2. Eye tracking

Visual fixations on preshift test trials replicated Experiment 1: Listeners took longer to preferentially fixate pictures for melodies that had the same AP information than those with different AP information (Fig. 6), and they took longer to fixate pictures for melodies that had matching relative-pitch information than those that had dissimilar relative-pitch information (Fig. 7). After the shift (Fig. 8), listeners were affected by a global change in pitch only on interference trials—that is, the trials where the AP of what they heard was a better match for the wrong answer than the right answer. Interestingly, listeners fixated the correct picture above chance before the unique interval of the melody, suggesting that global-relative pitch information was also used.

We first evaluated visual fixations to pictures on the pre-pitch-shift test trials. We selected two time windows before the unique intervals of paired melodies (which was $667 + 200 = 887$ ms: 200–550 and 550–900; after about 1,050 ms, the proportion of trials still ongoing began to drop rapidly, but these data are included in figures containing eye tracking data for Experiments 2 and 3 for the reader's information). An ANOVA on target advantage with AP Match, Pairing Type, and Time Window as factors confirmed our observations. Looks increased over time, effect of Time Window: $F_1(1, 35) = 58.32$, $p < .0001$; $F_2(1, 11) = 62.71$, $p < .0001$; $\eta^2_G = .12$. Dissimilar melodies were easier to distinguish than

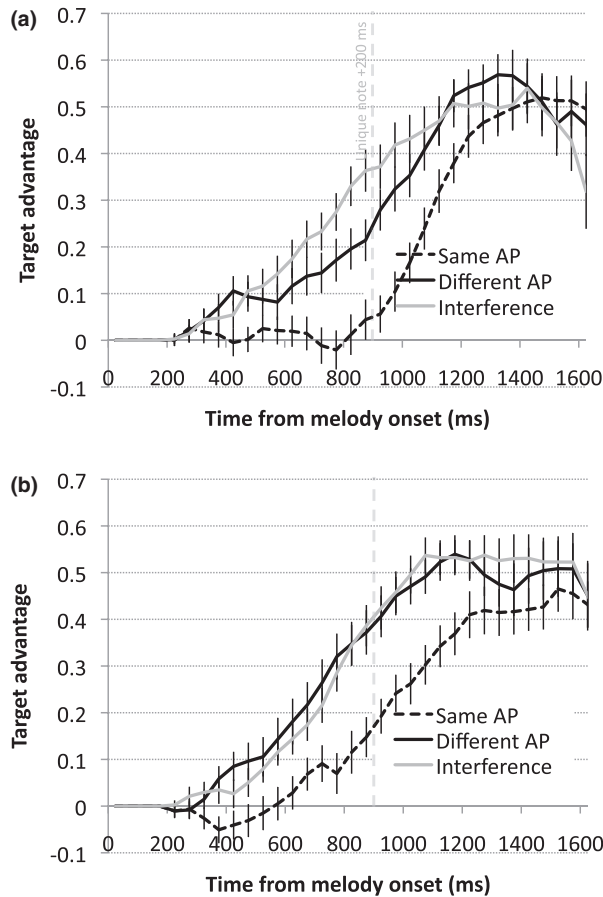


Fig. 6. Experiment 2 test trials, preshift. (a) Paired-melody trials and (b) dissimilar-melody trials. Error bars are standard errors.

paired melodies, at least in the later time window, interaction of Pairing Type \times Time Window: $F_1(1, 35) = 1.03$, $p = .003$; $F_2(1, 11) = 5.1$, $p < .05$; $\eta^2_G = .012$; 550–900 ms: $t_1(35) = 2.49$, $p = .02$; $t_2(11) = 1.96$, $p = .08$. Different-AP trials and interference trials, which did not differ, both showed more looks than same-AP trials in all windows, with the difference increasing across windows, AP Match: $F_1(2, 70) = 15.59$, $p < .0001$; $F_2(1, 11) = 10.72$, $p = .0006$; $\eta^2_G = .09$; Pairing Type \times Time Window interaction: $F_1(4, 140) = 9.68$, $p = .0002$; $F_2(2, 22) = 6.48$, $p = .006$; $\eta^2_G = .03$; 200–550: $p \leq .03$; 500–900: $p \leq .001$). No other effects reached significance.

We then evaluated the effects of pitch shift on looks (Fig. 8). The shift resulted in decrements in preferential target fixation from preshift to postshift trials, but only on *interference trials*, when the new (postshift) pitch of a melody matched the wrong shape's pitch better than the correct one's (Figs. 8c and 8f).

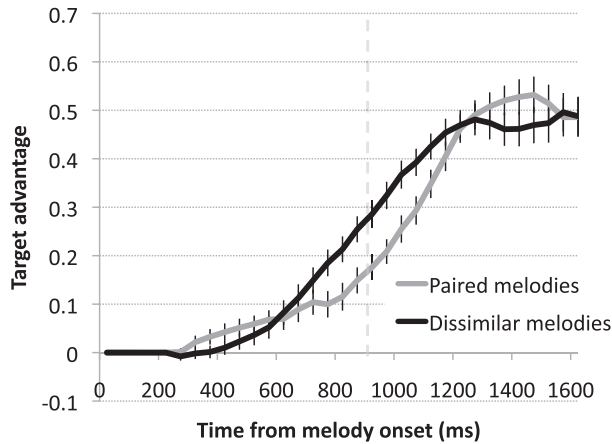


Fig. 7. Experiment 2: effect of relative-pitch similarity. Error bars are standard errors.

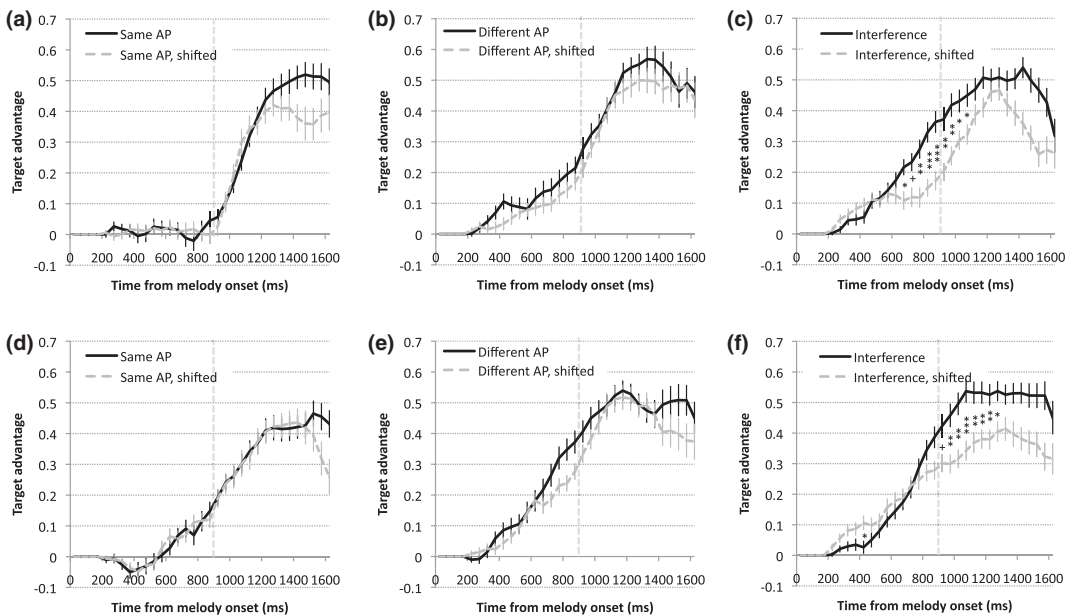


Fig. 8. Experiment 2: effects of global upward shift in pitch for each pitch-match condition. (a)–(c) are paired-melody trials, and (d)–(f) are dissimilar-melody trials. Error bars are standard errors.

We performed an ANOVA with Shift (unshifted or shifted), AP Match, Pairing Type, and Time Window as within-participants factors. Because we are interested primarily in the effect of globally changing the pitch level, we report only effects including Shift. Shifted trials did not show significantly lower target advantage overall, Shift effect: $F_1(1, 35) = 1.29$, $p = .26$; $F_2(1, 11) = 3.61$, $p = .08$; $\eta^2_G = .002$, but Shift did affect the rate of increase in

target advantage from the first to second time window, Shift \times Time Window interaction: $F_1(1, 35) = 4.7, p = .04$; $F_2(1, 11) = 6.37, p = .03$; $\eta^2_G = .004$. This drop patterned marginally differently for different pitch-match conditions, Shift \times AP Match \times Time Window interaction: $F_1(2, 70) = 2.85, p = .066$; $F_2(2, 22) = 2.98, p = .07$, $\eta^2_G = .004$. We examined the effects of Shift and Time Window for each AP Match condition separately, collapsing over Pairing Type, which did not interact ($F_s < 1$). There were no significant decrements in target advantage for same-AP or different-AP trials, but on interference trials, there was a Shift \times Time Window interaction: $F_1(1, 35) = 7.67, p = .009$; $F_2(1, 11) = 8.52, p = .01$; $\eta^2_G = .022$. This interaction resulted from a smaller gain in target advantage after the pitch shift, compared to before the pitch shift, from the first to second time window. To examine whether this result was carried by a small number of trials just after the pitch shift, we considered the Shift \times Time Window interaction difference score for only the second half of postshift trials and found it still significant, trials 49–96; $t_1(35) = 2.68, p = .01$; $t_2(11) = 2.7, p = .02$. Thus, this effect persists past trials just after the pitch shift.

A closer look at the interference trials for paired and dissimilar melodies suggests that the interference effect was carried by the paired trials. Significant differences in looking time between preshift and postshift trials are noted on Figs. 8c and 8f. The effect of AP interference appears earlier in time, and prior to the point where the melodies diverge in relative pitch, for the paired melodies. The effect is shifted later for unpaired melodies. This suggests that the effect for paired melodies is somewhat stronger. This might be the case because relative information (relative pitch, timing) is available to distinguish the dissimilar melodies.

An additional point is that, on paired interference trials, listeners show above-chance looks to the correct picture prior to the arrival of the final pitch interval, at 300–350 ms (Fig. 8c), $t_1(35) = 2.34, p = .03$; $t_2(11) = 1.77, p = .10$. If they were simply using local-relative pitch information and AP information to identify what melody they were hearing, this should not be possible. The fact that it happens demonstrates the role of global-relative pitch: Listeners have a sense of the pitch levels of the different melodies *in relation to each other*. Moreover, listeners used this information without instruction to do so, implying that they had encoded this global-relative information already.

3.3. Discussion

This experiment replicated and extended Experiment 1. Participants' eye movements reflected encoding of both AP as well as global-relative pitch. We maintained global-relative pitch cues while changing AP level by uniformly shifting all melodies up by six semitones. Test trials before the pitch shift replicated Experiment 1, with faster looks to shapes with different absolute or relative pitch information than shapes with the same information. Shifted test trials, which maintained global-relative pitch but altered AP, showed pitch interference: Looks to the correct shape declined specifically when the new AP values matched the wrong shape's melody better than the correct shape's melody.

We had asked whether the results of Experiment 1, showing that listeners used pitch information to recognize melodies online, could have been due in part to listeners encoding melodic pitch level relative to the pitch level of other melodies heard during the

experiment—global-relative pitch. The answer seems to be yes: Participants encode global-relative pitch information. Specifically, on postshift interference trials, participants fixate the correct picture above chance prior to the final note of the melody; they never show a *negative* target advantage, which they should if they only had AP information to go on. This suggests that the results of Experiment 1 (and preshift trials in Experiment 2) included influences of global-relative pitch.

These results also confirm that listeners encode absolute memory attributes that are dissociable from global-relative pitch encoding. That is, AP memory has an effect (interference) even when it becomes a misleading cue. Interference from AP mismatch emerges a bit more slowly than the benefit of global-relative pitch during the time course of the melody, though well before the two paired melodies diverge in local-relative -pitch information (Fig. 8c). This represents perhaps the clearest contrast between weighting AP information and relative pitch information: Both are used, but global-relative pitch shows up more immediately. The fact that the AP effects are not quite as swift suggests that listeners may recognize AP information about a melody more accurately when they have heard more of the AP range spanned by a given melody, rather than using APs of individual notes. Note that this explanation also applies to Schellenberg and Trehub (2003), who found good AP recognition of musical pieces that spanned wide pitch ranges both simultaneously and sequentially.

In sum, Experiment 2 found evidence for both global-relative pitch encoding and AP encoding. These results suggest that listeners do encode AP, and that its influence is separable from global-relative pitch. Nonetheless, listeners' facility with global-relative pitch information stands as an important methodological caution in future explorations of AP, as it can mimic effects of AP encoding. Given listeners' apparently ready apprehension of global-relative pitch, a final experiment explored the role of global-relative pitch in encoding melodies: Local-relative and global-relative information were pitted against each other to determine the strength of each as a factor in melody encoding.

4. Experiment 3

In Experiment 2, listeners were not only affected by a change in AP but also seemed to use global-relative pitch cues. How strong is this relative encoding of pitch range? To assess this question, a final experiment was devised which pitted absolute cues as well as local and global-relative pitch cues against each other. There were two conditions: a local cues condition and a local+global cues condition. In the local condition, there were no cues to recognition other than different notes (intervals). In the local+global cues condition, melodies were located across eight pitch ranges separated by three semitones. The local+global condition included a pitch shift which disrupted global-relative information.

If listeners distinguish melodies incrementally mostly based on local pitch or interval cues, then listeners in both local and local+global conditions should recognize melodies fastest when pitch cues diverge at interval 1, slightly slower at interval 2, and slowest at interval 4. If listeners use global-relative cues in addition to local pitch/interval information, then listeners in the local+global condition should recognize the melodies more rapidly than

those in the local condition. Finally, if listeners encode global-relative pitch range strongly, then local+global listeners should show more confusion when global-relative cues are disrupted, even if absolute cues stay the same.

4.1. Method

4.1.1. Participants

A total of 32 participants took part for course credit (29) or for pay (3). Fifteen additional participants took part but were excluded due to poor eye-tracking accuracy (10) or failure to learn (5, likely due to sleep loss). Two participants did not report music experience data. Among the rest, experience playing music ranged from 0 to 14 years ($M = 3.53$, $SD = 3.97$). Nine percent had had one or more music theory courses, 19% music history, and 63% had had no music coursework at all aside from performance groups.

4.1.2. Melodies

All melodies were generated in and exported from Finale using the whistle timbre, and were converted to .wav format and normalized to 70 dB in PRAAT. Melodies were five notes in length. For all melodies, the first four notes were each 125 ms in duration, and the last was 500 ms in duration (a tempo of 120 beats per minute). Thus, unlike Experiments 1 and 2, rhythm could not be used as a distinguishing cue, allowing us to verify that local-relative pitch alone has effects on recognition. In terms of relative pitch, melodies differed at the first interval (second note), the second interval (third note), or the fourth interval (fifth note).

There were two conditions: a condition with only local cues to melody, and a condition with both local and global cues to melody (henceforth “local” and “local+global”). In the local condition, all notes started on the same AP, meaning that listeners could not distinguish melodies until their intervals (and pitches) diverged. However, in the local+global condition, each melody started on a different AP, at a three-semitone spacing (details in Table 3). The keys of melodies in the local+global condition were mostly distantly related (three semitones apart or six semitones apart). Melodies differing at the fourth interval were always spaced exactly one octave apart. In this local+global condition, listeners should have ample pitch range cues to distinguish melodies, both in absolute terms and relative terms. Global-relative pitch was changed midway through the test (after a brief distractor break) by moving the upper (lower) four melodies down (or up) by two octaves, leaving the other four melodies at their original pitch levels (Fig. 9). Half of the local+global participants heard a higher range that shifted to a lower range, and the other half heard a lower range that shifted to a higher range.

4.1.3. Procedure

Listeners received training in 112-trial blocks. Order within a block was random. Each shape occurred as the target on 14 trials per block, and as the incorrect shape on 14 other trials. Every shape appeared with every other shape. Training proceeded until listeners achieved 90% correct or better within a block. They then completed two blocks of test trials. Again, order within a block was random. For listeners in the local condition, the two blocks

Table 3
Melodic stimuli used in Experiment 3








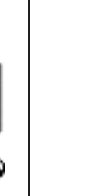
Melody	Local Condition				Global+Local Condition: Actual Starting Note								
	Pre- and Postshift					List 1		List 2		List 3		List 4 ^a	
	1	2	3	4		Hi ^b	Lo	Hi	Lo	Hi	Lo	Hi	Lo
	Eb4	C5	A5	F#6	C5	C5	F#5	F#5	Eb5	Eb5	A5	A5	A5
	Eb4	C5	A5	F#6	C6^c	C4^c	F#6	F#4	Eb6	Eb4	A6	A6	A4
	Eb4	C5	A5	F#6	Eb5	Eb5	C5	C5	A5	A5	F#5	F#5	F#5
	Eb4	C5	A5	F#6	Eb6	Eb4	C6	C4	A6	A4	F#6	F#6	F#4
	Eb4	C5	A5	F#6	F#5	F#5	A5	A5	C5	C5	Eb5	Eb5	Eb5
	Eb4	C5	A5	F#6	F#6	F#4	A6	A4	C6	C4	Eb6	Eb6	Eb4
	Eb4	C5	A5	F#6	A5	A5	Eb5	Eb5	F#5	F#5	C5	C5	C5
	Eb4	C5	A5	F#6	A6	A4	Eb6	Eb4	F#6	F#4	C6	C6	C4

Table 3
(Continued)

Example Trials Pictures Presented	Local (List 1)		Local+Global (List 1)	
	Pitch Distance (Semitone)	Intervals Diverge	Pitch Distance (Semitone)	Intervals Diverge
a1 versus a2	0	Fourth	12	Fourth
a1 versus b1	0	First	3	First
a1 versus b2	0	First	15	First
a1 versus c1	0	First	6	First
a1 versus c2	0	First	18	First
a1 versus d1	0	Second	9	Second
a1 versus d2	0	Second	21	Second
b1 versus a1 . . .	0	First	3	First

Notes. ^aLists 5–8 in local+global (not pictured) can be obtained by exchanging a1 & a2, b1 & b2, c1 & c2, and d1 & d2.

^bHalf of the listeners learned in a higher (Hi) range and were tested in that higher range and then a lower range (Lo). For the other half of participants, this was reversed.

^cFor local+global List 1, the melodies that changed in absolute pitch are in bold.

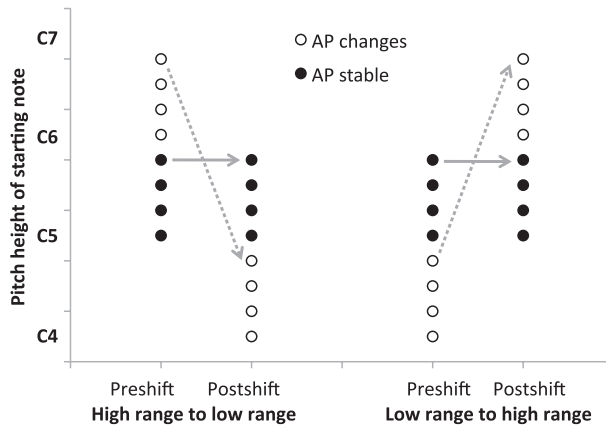


Fig. 9. Schematic of stimuli for Experiment 3 local+global condition, illustrating how the global-relative cues were changed from training and the first test block to the second test block. Each circle indicates the pitch level of the starting note of a melody. Half the melodies kept the same absolute pitch (filled circles), while the other half changed pitch by two octaves (hollow circles). Each gray arrow indicates the pitch movement of a particular melody. Pitch (y-axis) is in semitones, a log scale.

of test trials were identical. For listeners in the global condition, the second block shifted the pitches of four melodies. Notice that, both before and after the range-swap, melodies begin at eight different pitch levels at distances of three semitones. This means that if listeners are obligatorily attending to global-relative pitch—the pitch range *relative to the entire range of melodies they are hearing*—this change should disrupt recognition. For instance, the formerly highest melody will now be in the middle of the pitch range (first dashed arrow in Fig. 9). If they are attending to AP range, they should be more confused for the melodies which change pitch.

In between the first and second blocks of test trials, listeners were prompted to interact with the experimenter, who verbally administered a questionnaire on music-related issues. As in Experiment 2, this intervening discussion prevented listeners from engaging in any active rehearsal of pitch material. Results of this questionnaire were not analyzed. At the end of the experiment, participants completed a questionnaire about their experiences during the study.

4.2. Results

4.2.1. Accuracy

Listeners in the two different conditions learned in roughly the same amount of time (2.5 ± 0.97 blocks for local vs. 2.88 ± 0.89 for local+global) and were matched in accuracy for the first test block (Fig. 10). However, in the second test block, accuracy dropped markedly in the local+global condition.

An ANOVA with Condition (local, local+global) as a between-participants factor and Block (first, second) and Interval Divergence (interval 1, 2, or 4) as within-participants factors

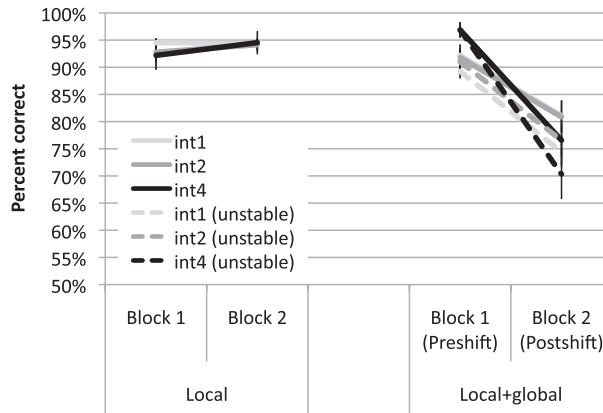


Fig. 10. Experiment 3, accuracy on pre-pitch-shift (Block 1) and postshift (Block 2) trials. Int(1,2,4) refers to the interval location where melodies differed in local-relative pitch. Error bars are standard errors.

confirmed these observations. Effects of Condition, $F_1(1, 30) = 18.66, p = .0002$; $F_2(1, 7) = 106.95, p < .0001$; $\eta^2_G = .23$, and Block, $F_1(1, 30) = 39.07, p < .0001$; $F_2(1, 7) = 68.75, p < .0001$; $\eta^2_G = .17$, were qualified by a Condition \times Block interaction, $F_1(1, 30) = 52.05, p < .0001$; $F_2(1, 7) = 102.59, p < .0001$; $\eta^2_G = .22$, and a Condition \times Block \times Interval Divergence interaction, $F_1(2, 30) = 5.47, p = .007$; $F_2(2, 14) = 6.85, p = .008$; $\eta^2_G = .03$. Due to this interaction, we analyzed effects of Block and Interval Divergence for each condition separately. For the local condition, effects of Block, $F_1(1, 15) = 1.87, p = .19$; $F_2(2, 14) = 1.31, p = .29, \eta^2_G = .008$, Interval Divergence ($F_s < 1$), and the interaction, $F_1 < 1, F_2(2, 14) = 1.08, p = .37$; $\eta^2_G = .006$, did not approach significance. This suggests that accuracy was equivalent on trials where melodies diverged early versus late, and that simply testing for a second block did not decrease accuracy.

For the local+global condition, however, the pattern differed. In this analysis, we included an additional factor, Stability, which indexed whether the target melody was one that changed AP or stayed at the same AP from test Block 1 to test Block 2. There was an effect of Block, $F_1(1, 15) = 51.77, p < .0001$; $F_2(1, 7) = 125.72, p < .0001$; $\eta^2_G = .33$, with accuracy decreasing from Block 1 (pre-pitch shift) to Block 2 (post-pitch shift). There was also a Block \times Interval Divergence interaction, $F_1(2, 30) = 6.15, p = .006$; $F_2(2, 14) = 6.15, p = .01$; $\eta^2_G = .044$, with a larger decrement in accuracy for melodies that diverged late in local-relative pitch cues. Nonetheless, all showed significant drops in accuracy, $t_1(15) \geq 4.82, p \leq .0002$; $t_2(11) \geq 6.2, p \leq .0004$. This pattern of results—a bigger accuracy drop for melodies that diverged late—likely results from the pitch separations in the late-diverging melodies, which were on average larger than for the early-diverging melodies. The Block \times Stability interaction did not reach significance, $F_1(1, 15) = 2.49, p = .14$; $F_2(1, 7) = 2.2, p = .18$; $\eta^2_G = .09$, although the means were consistent with a larger drop in accuracy on postshift trials where the melody had changed in AP, and postshift trials with stable-AP target melodies were more accurate than trials with changed-AP melodies, $t_1(15) = 2.14, p < .05$; $t_2(7) = 2.42, p < .05$.

4.2.2. Eye tracking

Figure 11 suggests that eye-tracking results paralleled the accuracy data. Recognition speed based on local cues (Fig. 11a) was incremental in the pattern expected by the time points where melodies' intervals diverged. Second, global-relative information, when intact, afforded an advantage in recognition speed over local cues (Fig. 11a, solid lines, vs. Fig. 11b, solid lines).

An ANOVA over Condition \times Block \times Interval Divergence \times Time Window (200–450, 450–700) on the target advantage verified these observations. There was a Condition \times Block interaction, $F_1(1, 30) = 20.51, p < .0001$; $F_2(1, 7) = 15.29, p = .006$; $\eta^2_G = .07$, such that the local+global condition showed higher target advantage than the local condition in the first block, $t_1(30) = 5.87, p < .0001$; $t_2(7) = 4.96, p = .002$, but not in the second block,

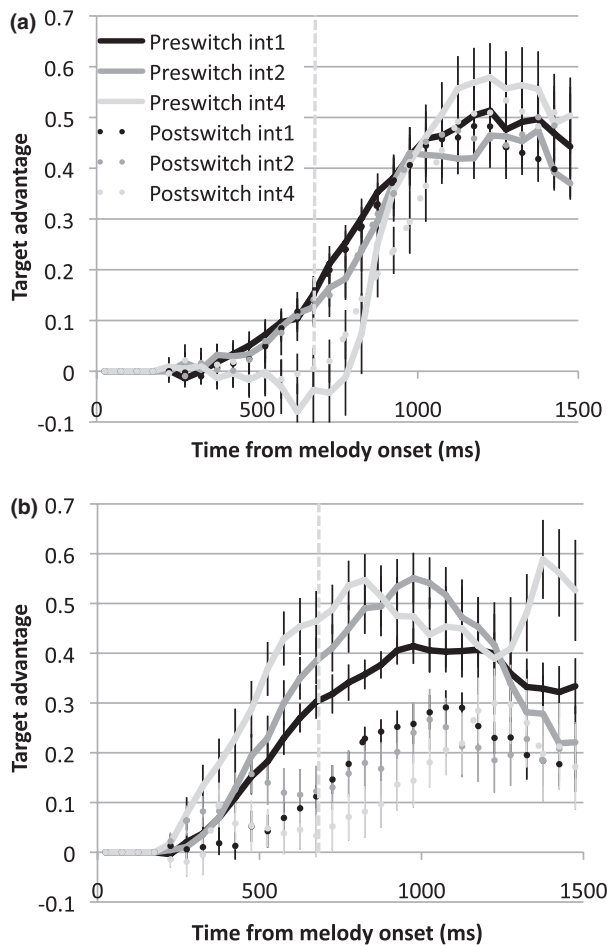


Fig. 11. Experiment 3, target advantage across time for (a) the local cues condition and (b) the local+global cues condition. Int(1,2,4) refers to the interval location where melodies differed in local-relative pitch. Error bars are standard errors.

$t_1(30) = 1.42, p = .17$; $t_2(7) = 1.44, p = .19$. This implies that additional global-relative cues enhanced recognition speed while global cues were valid but did not when global-relative cues became invalid. At the highest level, there were Condition \times Block \times Time Window, $F_1(1, 30) = 42.74, p < .0001$; $F_2(1, 7) = 29.74, p < .001$; $\eta^2_G = .04$, and Condition \times Interval Divergence \times Time Window, $F_1(2, 60) = 5.17, p = .009$; $F_2(2, 14) = 3.9, p = .04$; $\eta^2_G = .03$, interactions. To address these interactions, and to verify hypotheses about the effects of changes in global-relative pitch, we analyzed each condition separately.

For the local condition, looking patterns did not change from Block 1 to Block 2. There was an effect of Interval Divergence, $F_1(2, 30) = 3.52, p = .04$; $F_2(2, 14) = 5.12, p = .02$; $\eta^2_G = .05$, an effect of Time Window, $F_1(1, 15) = 16.3, p = .001$; $F_2(1, 7) = 7.83, p = .03$; $\eta^2_G = .04$, and an interaction of the two, $F_1(2, 30) = 8.26, p = .001$; $F_2(2, 14) = 8.56, p = .004$; $\eta^2_G = .05$. This pattern indicated that looks to items that diverged at the first or second interval increased from the first time window to the second time window, first: $t_1(15) = 6.7, p < .0001$; $t_2(7) = 3.33, p = .01$; second: $t_1(15) = 3.53, p = .003$; $t_2(7) = 4.1, p = .005$, while looks to items that diverged at the fourth interval did not increase, $t_1(15) = 1.02, p = .32$; $t_2(7) = 1.05, p = .33$; although they exceeded chance *after* this point. This is consistent with gradient looking patterns depending on the point where the melodies diverged. There were no effects of Block (all $F_s < 1$), suggesting that simply testing for a second block of trials did not change patterns of looking.

For the local+global condition, as with the accuracy data, we included Stability (whether a melody changed AP level from the first to second test block) as a factor. The effect of Block was significant, $F_1(1, 15) = 34.91, p < .0001$; $F_2(1, 7) = 39.39, p = .0004$; $\eta^2_G = .12$, indicating a drop in target advantage after the pitch shift. Like the accuracy data, this is consistent with confusion based on the change in global-relative cues. Second, the effect of Time Window, $F_1(1, 15) = 60.1, p < .0001$; $F_2(1, 7) = 66.1, p < .0001$; $\eta^2_G = .13$, was qualified by a Block \times Time Window interaction, $F_1(1, 15) = 85.98, p < .0001$; $F_2(1, 7) = 71.26, p < .0001$; $\eta^2_G = .06$. This interaction suggested that an increase in target advantage from the first to second time window was smaller after the change in global-relative cues. Finally, there was an effect of Block \times Interval Divergence, $F_1(2, 30) = 4.65, p = .02$; $F_2(2, 14) = 7.9, p = .005$; $\eta^2_G = .038$. This resulted from a larger postshift drop in target advantage for the melodies diverging at the fourth interval than for those diverging at the first or second interval, though all were significant, $t_1(15) \geq 5.38, p < .0001$; $t_2 \leq 5.03, p \leq .002$. The Block \times Stability interaction did not reach significance ($F_1 < 1, F_2 < 1$), suggesting that the effect of changing global-relative cues did not affect looks to targets where the melody changed AP any more than to targets where the melody's AP remained the same.

4.2.3. Information used in local+global condition

We wanted to examine more closely how listeners were using relative range information to identify the melodies in the local+global condition (before the change in global cues). Although they must be using local cues to some extent to attain above-chance recognition, two types of global-relative information seemed to affect recognition: how close to the edge of the overall pitch range each melody was, and how far apart the starting notes of the two melodies on a trial were. In fact, the pitch distance between starting notes explains the

preshift advantage in accuracy for melodies differing at the fourth interval: These melodies were always separated by 12 semitones, whereas melodies differing at the first and second intervals were separated by varying pitch differences, which were usually < 12 semitones because there were simply fewer large-pitch-distance pairings than small-pitch-distance pairings. The edge-proximity advantage echoes classic work by Pollack (1952) on listeners' abilities to discriminate pitch categories, which was most accurate for tones at the edges of pitch ranges.

To quantify the effects of within-trial pitch differences and edge-of-range effects on accuracy (Fig. 12a), we conducted a regression analysis with Pitch Difference (first notes were

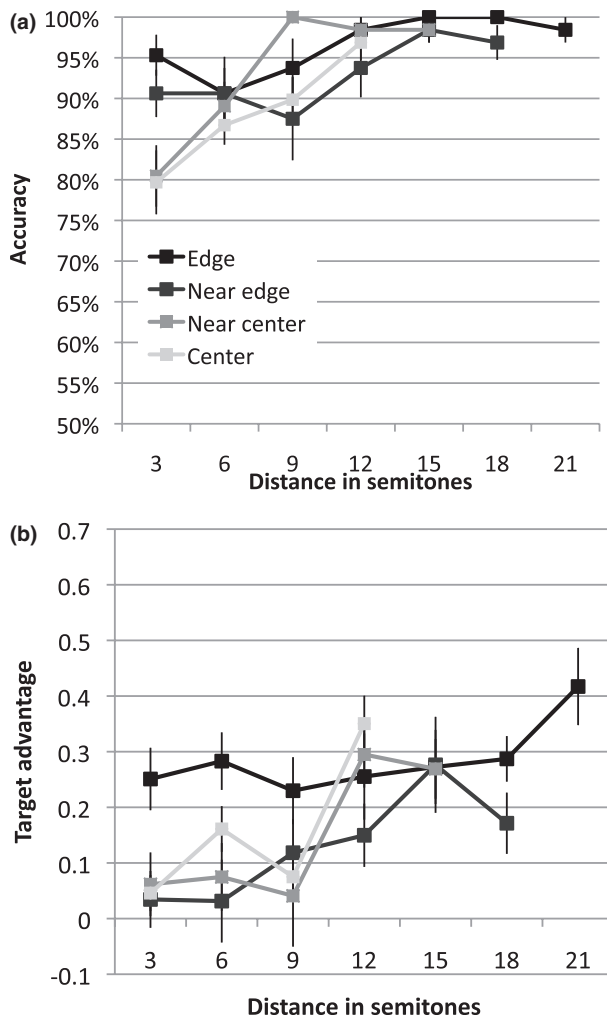


Fig. 12. Experiment 3, accuracy (a) and target advantage from 200 to 700 ms (b) on preshift test trials, displayed by the pitch distance between the initial notes of the two pictures' melodies (x-axis) and the proximity of the target melody to the edge of the pitch range (black = edge, to light gray = central).

3, 6, 9, 12 semitones apart) and Edge Proximity (on edge of pitch range, near edge, near center, center) as factors. Pitch differences larger than 12 semitones were omitted from analysis, because for the most central melodies, the maximum pitch difference was 12 semitones. There was an effect of Pitch Difference, $F_1(1, 15) = 22.4$, $p = .0003$; $F_2(1, 7) = 28.53$, $p = .001$; $\eta^2_G = .23$, indicating that larger pitch differences led to higher accuracy. There was also an effect of Edge Proximity, $F_1(1, 15) = 6.37$, $p = .02$; $F_2(1, 7) = 51.39$, $p = .0002$; $\eta^2_G = .07$, showing that melodies close to the edges of the global pitch range were identified more accurately. Finally, there was a Pitch Difference \times Edge Proximity interaction, $F_1(1, 15) = 11.07$, $p = .005$; $F_2(1, 7) = 12.31$, $p < .01$; $\eta^2_G = .10$. The interaction indicated that listeners were more accurate across the board for edgemost melodies but were affected more by pitch differences when the target melody was closer to the center of the pitch range (although simple effects of pitch difference at each level of Edge Proximity did not reach significance).

A similar regression was conducted on target advantage (depicted in Fig. 12b) to assess effects of pitch difference and edge proximity. To simplify analyses, we collapsed across the two time windows. As with accuracy, there was an effect of Pitch Difference, $F_1(1, 15) = 13.43$, $p = .002$; $F_2(1, 7) = 25.63$, $p = .001$; $\eta^2_G = .016$, and a marginal effect of Edge Proximity, $F_1(1, 15) = 4.54$, $p = .05$; $F_2(1, 7) = 2.96$, $p = .13$; $\eta^2_G = .005$. These indicated that listeners showed larger target advantage when the pitch difference between target and distractor melodies was larger, and (marginally) when melodies were closer to the edge of the pitch range. The interaction of Pitch Difference \times Edge Proximity was significant, $F_1(9, 135) = 10.41$, $p = .006$; $F_2(1, 7) = 22.56$, $p = .002$; $\eta^2_G = .007$. The interaction implies that listeners had overall higher target advantage for melodies at the edge of the pitch range (no effect of Pitch Difference for the edgemost melodies, $F_s < 1$), but they were more affected by pitch differences for melodies closer to the center of the pitch range, near edge: $F_1(1, 15) = 8.27$, $p = .01$; $F_2(1, 7) = 9.62$, $p = .02$; $\eta^2_G = .014$; near center: $F_1(1, 15) = 8.73$, $p < .01$; $F_2(1, 7) = 43.1$, $p = .0003$; $\eta^2_G = .028$; center: $F_1(1, 15) = 21.67$, $p = .0003$; $F_2(1, 7) = 13.49$, $p = .008$; $\eta^2_G = .061$. More generally, these patterns of results support the hypothesis that listeners were utilizing global-relative pitch information to distinguish the melodies.

4.3. Discussion

In this final experiment, we examined the strength of global-relative pitch cues by presenting global cues and then disrupting global-relative pitch information. Global-relative pitch appears to be a strong cue for identification: Recognition was more rapid when global-relative pitch cues were present, and disruption of global-relative cues led to large decrements in accuracy and visual fixations. Note that this cannot be explained away by saying that the AP change was so drastic that recognition was impaired, because listeners experienced a similar recognition deficit for melodies whose APs had not changed. Earlier researchers went to great pains to eliminate a relative-pitch “solution” for tests of AP perception, such as testing individuals on only a single note per day, to eliminate

context-relative effects (Petran, 1932; see Ward, 1999, for a review). Our results suggest that this extreme respect for human relative pitch encoding was warranted.

As an additional check on the role of pitch versus rhythm in distinguishing melodies in earlier experiments, we also constrained differences in melodies to the pitch dimension, with rhythm identical across all melodies. Despite the absence of rhythm differences, identification was good, and identification based on local pitch cues alone (local condition) happened roughly as soon as melodies diverged in pitch. Overall, the results accord with the previous experiments in showing incremental recognition, and additionally suggest a very strong role for global-relative pitch in memory encoding.

5. General discussion

Three experiments examined whether, and how, humans identify melodies online. In each experiment, listeners learned melody-shape associations, and we monitored their eye movements to shape alternatives. By manipulating similarity of the heard melody to the incorrect shape's melody, we obtained information about how listeners were weighting absolute and relative pitch cues during recognition. We found that, with a relatively brief exposure in the lab, listeners used pitch differences to distinguish melodies online (Experiments 1 and 2). These results did not address, though, whether participants were encoding pitch range information absolutely or in relative terms, or both. Experiment 2 maintained global-relative pitch information while changing AP similarity. The results suggested that both absolute and global-relative information were at work. A change in AP cues caused interference (attenuated visual fixations), though listeners were still able to discriminate melodies early in the melody, suggesting that they were using global-relative pitch information as well as absolute encoding of pitch. While Experiment 2 held global-relative information constant, Experiment 3 investigated the strength of global-relative pitch information by perturbing global-relative pitch. This manipulation strongly impeded recognition even for melodies that maintained their AP levels, suggesting that listeners are highly dependent on global-relative pitch information in encoding melodies.

At the outset, we wanted to discover whether music recognition, like word recognition, was incremental in nature. This appears to be the case. Though incremental recognition itself is perhaps not surprising, it has not been definitively demonstrated before, nor has the set of cues that lead to rapid recognition been assessed (although Schellenberg et al., 1999; suggest timbre as a likely information source). Our work suggests that several types of pitch information, including AP, lead to recognition in under 1 s. As noted in the Introduction, use of AP information is not a given: Eye movements do not reflect recognition when listeners are not natively sensitive to a phoneme (Weber & Cutler, 2004). However, our listeners do use AP information. Our results do not mean that listeners can always identify music from brief excerpts, but that they readily do so in the absence of instructions to that effect. This verifies that the short-interval identification results of Schellenberg et al. (1999) may happen routinely as listeners experience music in real time. Moreover, this connects music processing to a rich literature in online recognition of spoken language. It suggests that

incremental processing is a natural property of temporal (or at least temporal auditory) events, and that pressure to communicate rapidly is not required. Of course, incremental processing in music may be related to incremental processing in language in that they may share the same evolutionary substrates (e.g., Patel, 2008).

Our results also suggest that relative pitch is encoded at a global scale, separately from AP information, and is extremely important to recognition. This has implications for relative and absolute recognition in a variety of domains, in suggesting that multiple types of representations can coexist and compete with each other, rather than simple variation in the degree of absolute-recognition and relative-recognition acuity. Finally, our results suggest that pitch content alone (Experiment 3, local condition) is sufficient for online recognition in the absence of rhythm differences. It is not possible to determine from our results what sort of local-relative pitch information—interval, contour, or scale-step—is most efficacious in distinguishing melodies in real time. This is partly because we deliberately covaried these cues to make melodies as discriminable as possible, so that we were not compelled to use a musically elite sample of participants. Nonetheless, it remains an interesting topic for future research. For instance, are listeners sensitive to changes in interval which do not change contour? Do listeners more rapidly distinguish melodies if they differ on a feature that indicates they belong to different keys (e.g., D–E–C–F, which is likely in C major or a related key, vs. D–E–C#–F#, which is likely in A major or a related key) rather than the same key (D–E–C–F vs. D–E–B–G, which could both be in C major)? The visual fixation technique used in the current study can be used in combination with measures of accuracy to explore such questions.

We also wanted to know how adults combine absolute and relative pitch information in recognition. The answer is that AP is slightly slower, compared to global-relative pitch. In Experiment 2, the cleanest comparison between absolute and relative cues, we rendered AP cues invalid by shifting all melodies upward by a fixed pitch distance, while maintaining global relative-pitch relationships. This meant that a subset of melodies would now occur at the old AP level of the incorrect-response melodies. This pitted global-relative cues against absolute cues. In this unreinforced postshift phase of the experiment, listeners made slightly more errors and showed marked decreases in looking preference to the target when AP interference was present. Impressively, listeners also showed evidence of recognizing the correct melody early in the trial, which could only have been accomplished by global-relative pitch information; AP-range effects showed up slightly later in the time course of the melody. This pattern of results suggests that (at least in our adult sample) absolute encoding may be weaker than global-relative pitch encoding, given its slow emergence on interference trials in Experiment 2, as well as its relatively meager effect when global-relative cues were disrupted in Experiment 3. Another explanation for the slow emergence of AP information in Experiment 2 is that it is more robust when the stimulus is richer than one note (an explanation that may also hold for earlier demonstrations of implicit AP such as Schellenberg & Trehub, 2003).

Global-relative pitch information may also be linked to use of relative pitch in language, consistent with theorizing by Saffran and Griepentrog (2001) and Takeuchi and Hulse (1993). For instance, in English, a pitch rise from 100 to 120 Hz across the course of a

sentence means the same thing as a pitch rise from 180 to 216 Hz—that the speaker is asking a question. That is, to understand English prosody, a listener must be able to recognize the identity between two patterns that mismatch in AP but match in relative pitch. This language-driven learning might then push listeners into a relative frame of processing for music as well. This provides a relative-pitch counterpart to Deutsch, Henthorn, and Dolson's (2004; Henthorn & Deutsch, 2007) theorizing that AP information in language processing may influence musical AP perception.

A final point worth noting is that adults' overt responses were overwhelmingly based on relative pitch, even after the pitch shift in Experiment 2, where no reinforcement was provided. This suggests that relative pitch, even if it becomes evident at a (slightly) later time point, is a more valid cue for adult listeners. Despite the pressure of training and adults' apparent tendency to respond relatively, eye movements are affected by AP information, verifying the sensitivity of the paradigm. If adults, who are the paradigmatic case for relative processing, show some sensitivity to absolute identification, then presumably any population that weighs absolute information more strongly will show even more robust effects.

We had endeavored to develop a paradigm that assayed music recognition without requiring an overt response, and that showed sensitivity to both relative and AP processing. In that, we appear to have succeeded. Adults' visual fixations to melody-associated pictures reflect the activation of both relative pitch information and AP information. Presumably, this is linked to listeners' weighting of relative and absolute cues to melody identity. In word recognition, listeners' weightings of particular pieces of information (e.g., voice onset time and vowel duration) influence the likelihood of eye movements toward particular targets (e.g., McMurray et al., 2008). It is in the relative weighting of these different cues that one might expect to find evidence of an absolute-to-relative processing shift.

This study establishes a benchmark for pitch cue weighting in rapid adult recognition of melodies, and it highlights several important considerations in examining pitch perception (the role of global-relative pitch). This benchmark sets the stage for extending the paradigm to different age groups, allowing us to assess whether there is a developmental shift in the weighting of absolute versus relative cues in music perception. Even given the same set of melodies to learn, children may encode the melodies differently than adults would. On the other hand, young humans, along with other mammals, may be predisposed to encode pitch in more relative terms from the outset (Weisman et al., 2004, 2006). With a paradigm which can detect differences in cue weighting and can be applied across a wide age range, we are a step closer to understanding whether these encoding differences truly exist.

6. Conclusion

We have presented results suggesting use of multiple types of pitch memory in rapid online recognition. Adult listeners use AP memory, as well as both local and global relative-pitch memory, in recognition. More generally, we have demonstrated incremental processing in a nonlinguistic domain, suggesting that incrementality is the norm for recognition of

temporal stimuli, rather than the result of pressure to comprehend language rapidly. Finally, we have established a paradigm that can be used to explore postulated developmental shifts in pitch processing.

Acknowledgments

In fond memory of Dorothy K. Payne (1935–2010), music theorist, pianist, author, and mentor. Thanks to Heather Pelton, Russ Fluty, and Alicia Zuniga for running participants, and to Ani Patel for helpful comments on a previous version of this manuscript.

References

- Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition: Evidence for continuous mapping models. *Journal of Memory and Language*, 38, 419–439.
- Bakeman, R. (2005). Recommended effect size statistics for repeated measures designs. *Behavior Research Methods*, 37, 379–384.
- Bartlett, J. C., & Dowling, W. J. (1980). Recognition of transposed melodies: A key-distance effect in developmental perspective. *Journal of Experimental Psychology: Human Perception and Performance*, 6, 501–515.
- Ben-Haim, M. S., Chajut, E., & Eitan, Z. (2010). *Common and rare musical keys are absolutely different: Implicit absolute pitch, exposure effects, and pitch processing*. Talk presented at the 11th International Conference on Music Perception and Cognition, Seattle, WA.
- Boersma, P., & Weenink, D. (2009). Praat: Doing phonetics by computer [Computer program]. Version 5.1.43. Available at: <http://www.praat.org/>. Accessed October 30, 2009.
- Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, 10, 433–436.
- Chang, H. W., & Trehub, S. E. (1977). Auditory processing of relational information by young infants. *Journal of Experimental Child Psychology*, 24, 324–331.
- Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language. A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology*, 6, 84–107.
- Cornelissen, F. W., Peters, E., & Palmer, J. (2002). The Eyelink Toolbox: Eye tracking with MATLAB and the Psychophysics Toolbox. *Behavior Research Methods, Instruments & Computers*, 34, 613–617.
- Creel, S. C., Aslin, R. N., & Tanenhaus, M. K. (2006). Acquiring an artificial lexicon: Effects of segment type and order information. *Journal of Memory and Language*, 54, 1–19.
- Creel, S. C., Aslin, R. N., & Tanenhaus, M. K. (2008). Heading the voice of experience: The role of talker variation in lexical access. *Cognition*, 108, 633–664.
- Creel, S. C., Tanenhaus, M. K., & Aslin, R. N. (2006). Consequences of lexical stress on learning an artificial lexicon. *Journal of Experimental Psychology: Learning, Memory & Cognition*, 32, 15–32.
- Deutsch, D., Henthorn, T., & Dolson, R. M. (2004). Absolute pitch, speech, and tone language: Some experiments and a proposed framework. *Music Perception*, 21, 339–356.
- Dowling, W. J., & Fujitani, D. S. (1971). Contour, interval, and pitch recognition in memory for melodies. *Journal of the Acoustical Society of America*, 49, 524–531.
- Fiser, J., & Aslin, R. N. (2005). Encoding multi-element scenes: Statistical learning of visual feature hierarchies. *Journal of Experimental Psychology: General*, 134, 521–537.
- Gjerdingen, R., & Perrott, D. (2008). Scanning the dial: The rapid recognition of music genres. *Journal of New Music Research*, 37, 93–100.

- Hallett, P. E. (1986). Eye movements. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance* (pp. 10:1–10:112). New York: Wiley.
- Halpern, A. R. (1989). Memory for the absolute pitch of familiar songs. *Memory & Cognition*, 17, 572–581.
- Henthorn, T., & Deutsch, D. (2007). Ethnicity versus early environment: Comment on “Early childhood music education and predisposition to absolute pitch: Teasing apart genes and environment.” *American Journal of Medical Genetics*, 143A, 102–103.
- Hulse, S. H., Cynx, J., & Humpal, J. (1984). Absolute and relative pitch discrimination in serial pitch perception by birds. *Journal of Experimental Psychology: General*, 113, 38–54.
- Justus, T., & List, A. (2005). Auditory attention to frequency and time: An analogy to visual local—Global stimuli. *Cognition*, 98, 31–51.
- Keenan, J. P., Halpern, A. R., Thangaraj, V., Chen, C., Edelman, R. R., & Schlaug, G. (2001). Absolute pitch and planum temporale. *Neuroimage*, 14, 1402–1408.
- Krumhansl, C. L. (2010). Thin slices of music. *Music Perception*, 27, 337–354.
- Krumhansl, C. L., & Keil, F. C. (1982). Acquisition of the hierarchy of tonal functions in music. *Memory & Cognition*, 10, 243–251.
- Levitin, D. J. (1994). Absolute memory for musical pitch: Evidence from the production of learned melodies. *Perception and Psychophysics*, 56, 414–423.
- Lynch, M. P., Eilers, R. E., Oller, D. K., & Urbano, R. C. (1990). Innateness, experience, and music perception. *Psychological Science*, 1, 272–276.
- MacDougall-Shackleton, S. A., & Hulse, S. H. (1996). Concurrent absolute and relative pitch processing by European starlings (*Sturnus vulgaris*). *Journal of Comparative Psychology*, 110, 139–146.
- Magnuson, J. S., Tanenhaus, M. K., Aslin, R. N., & Dahan, D. (2003). The time course of spoken word learning and recognition: Studies with artificial lexicons. *Journal of Experimental Psychology: General*, 132, 202–227.
- McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, 18, 1–86.
- McMurray, B., & Aslin, R. N. (2004). Anticipatory eye movements reveal infants’ auditory and visual categories. *Infancy*, 6, 203–229.
- McMurray, B., Clayards, M. A., Tanenhaus, M. K., & Aslin, R. N. (2008). Tracking the timecourse of phonetic cue integration during spoken word recognition. *Psychonomic Bulletin and Review*, 15, 1064–1071.
- Meyer, L. B. (1967). *Music, the arts, and ideas; Patterns and predictions in twentieth-century culture*. Chicago: U Chicago Press.
- Mirman, D., & Magnuson, J. S. (2009). Dynamics of activation of semantically similar concepts during spoken word recognition. *Memory & Cognition*, 37(7), 1026–1039.
- Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology*, 9, 353–383.
- Olejnik, S., & Algina, J. (2003). Generalized eta and omega squared statistics: Measures of effect size for some common research designs. *Psychological Methods*, 8, 434–447.
- Palmer, C., Jungers, M. K., & Jusczyk, P. W. (2001). Episodic memory for musical prosody. *Journal of Memory and Language*, 45, 526–545.
- Patel, A. D. (2008). *Music, language, and the brain*. New York: Oxford University Press.
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, 10, 437–442.
- Petran, L. A. (1932). An experimental study of pitch recognition. *Psychological Monographs*, 42(6), 1–120.
- Pollack, I. (1952). The information of elementary auditory displays. *Journal of the Acoustical Society of America*, 24(6), 745–749.
- Raaijmakers, J. G. W. (2003). A further look at the “Language-as-Fixed-Effect Fallacy.” *Canadian Journal of Experimental Psychology*, 57, 141–151.

- Raaijmakers, J. G. W., Schrijnemakers, J. M. C., & Gremmen, F. (1999). How to deal with "The Language-as-Fixed-Effect Fallacy": Common misconceptions and alternative solutions. *Journal of Memory and Language*, 41, 416–426.
- Saffran, J. R. (2003). Absolute pitch in infancy and adulthood: The role of tonal structure. *Developmental Science*, 6, 37–49.
- Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, 274, 1926–1928.
- Saffran, J. R., & Griepentrog, G. J. (2001). Absolute pitch in infant auditory learning: Evidence for developmental reorganization. *Developmental Psychology*, 37, 74–85.
- Saffran, J. R., Reeck, K., Niebuhr, A., & Wilson, D. (2005). Changing the tune: The structure of the input affects infants' use of absolute and relative pitch. *Developmental Science*, 8, 1–7.
- Schellenberg, E. G., Iverson, P., & McKinnon, M. C. (1999). Name that tune: Identifying popular recordings from brief excerpts. *Psychonomic Bulletin & Review*, 6, 641–646.
- Schellenberg, E. G., & Trehub, S. E. (2003). Good pitch memory is widespread. *Psychological Science*, 14, 262–266.
- Sergeant, D. (1969). Experimental investigation of absolute pitch. *Journal of Research in Music Education*, 17, 135–143.
- Sergeant, D., & Roche, S. (1973). Perceptual shifts in the auditory information processing of young children. *Psychology of Music*, 1, 39–48.
- Smith, N. A., & Schmuckler, M. A. (2008). Dial A440 for absolute pitch: Absolute pitch memory by non-absolute pitch possessors. *Journal of the Acoustical Society of America*, 123, EL77–EL84.
- Spivey, M. J., Grosjean, M., & Knoblich, G. (2005). Continuous attraction toward phonological competitors. *Proceedings of the National Academy of Sciences*, 102(29), 10393–10398.
- Swingle, D., Pinto, J. P., & Fernald, A. (1999). Continuous processing in word recognition at 24 months. *Cognition*, 71, 73–108.
- Takeuchi, A. H., & Hulse, S. H. (1993). Absolute pitch. *Psychological Bulletin*, 113, 345–361.
- Tanenhaus, M., Spivey-Knowlton, M., Eberhard, K., & Sedivy, J. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, 268, 1632–1634.
- Taylor, P. (1997). Barfly [Computer program]. Version 1.3b1. Available at: <http://www.barfly.dial.pipex.com/>. Accessed September 25, 2007.
- Toiviainen, P., & Krumhansl, C. L. (2003). Measuring and modeling real-time responses to music: The dynamics of tonality induction. *Perception*, 32, 741–766.
- Trehub, S. E., Schellenberg, E. G., & Kamenetsky, S. B. (1999). Infants' and adults' perception of scale structure. *Journal of Experimental Psychology: Human Perception and Performance*, 25, 965–975.
- Trehub, S. E., Schellenberg, E. G., & Nakata, T. (2008). Cross-cultural perspectives on pitch memory. *Journal of Experimental Child Psychology*, 100, 40–52.
- Truitt, F. E., Clifton, C., Pollatsek, A., & Rayner, K. (1997). The perceptual span and the eye-hand span in sight reading music. *Visual Cognition*, 4, 143–161.
- Ward, W. D. (1999). Absolute pitch. In D. Deutsch (Ed.), *Psychology of music* (2nd ed., pp. 265–298). New York: Academic Press.
- Weber, A., & Cutler, A. (2004). Lexical competition in non-native spoken-word recognition. *Journal of Memory and Language*, 50, 1–25.
- Weisman, R. G., Njegovan, M., Williams, M. T., Cohen, J., & Sturdy, C. B. (2004). A behavior analysis of absolute pitch: Sex, experience, and species. *Behavioural Processes*, 66, 289–307.
- Weisman, R. G., Williams, M. T., Cohen, J. S., Njegovan, M. J., & Sturdy, C. B. (2006). The comparative psychology of absolute pitch. In E. A. Wasserman & T. R. Zentall (Eds.), *Comparative cognition: Experimental explorations of animal intelligence* (pp. 71–86). New York: Oxford.