Sing that tune: Infants' perception of melody and lyrics and the facilitation of phonetic recognition in songs

Gina C. Lebedeva *, Patricia K. Kuhl

Institute for Learning and Brain Sciences, Department of Speech and Hearing Sciences, University of Washington, United States

A B S T R A C T

To better understand how infants process complex auditory input, this study investigated whether 11-month-old infants perceive the pitch (melodic) or the phonetic (lyric) components within songs as more salient, and whether melody facilitates phonetic recognition. Using a preferential looking paradigm, uni-dimensional and multi-dimensional songs were tested; either the pitch or syllable order of the stimuli varied. As a group, infants detected a change in pitch order in a 4-note sequence when the syllables were redundant (experiment 1), but did not detect the identical pitch change with variegated syllables (experiment 2). Infants were better able to detect a change in syllable order in a sung sequence (experiment 2) than the identical syllable change in a spoken sequence (experiment 1). These results suggest that by 11 months, infants cannot "ignore" phonetic information in the context of perceptually salient pitch variation. Moreover, the increased phonetic recognition in song contexts mirrors findings that demonstrate advantages of infant-directed speech. Findings are discussed in terms of how stimulus complexity interacts with the perception of sung speech in infancy.

© 2010 Elsevier Inc. All rights reserved.

1. Sing that tune: infants' perception of melody and lyrics and in songs

Pre-verbal infants are able to attend to both pitch and phonetic patterns in complex auditory speech signals (see Gerken & Aslin, 2005; McMullen & Saffran, 2004). Research directed at the interaction between the two can test to what extent infants can selectively attend to one particular auditory dimension within a multi-dimensional signal, and how this ability might be influenced by linguistic or auditory experience. Lyrical songs, as in words sung to a tune, offer an appealing and natural multi-dimensional context to investigate both segmental (phonetic) and suprasegmental (melodic and/or rhythmic) pattern perception. Songs are composed of two primary dimensions: music and language (in the form of speech), each containing a series of hierarchically structured events. How humans have come to evolve, learn, and use music and speech has produced several lines of research that investigate how music and language processing interact (Patel, 2003). Despite the increasing number of empirical investigations in adults on this topic, data from infants would be valuable in helping understand not only what kinds of learning mechanisms are unique to language or general to cognition (see McMullen & Saffran, 2004), but also what types of acoustic contexts facilitate language acquisition (e.g., Chen-Hafteck, 1997; Schön, Magne, & Besson, 2004; Trainor & Desjardins, 2002).

Though musical processing has customarily been associated with the right hemisphere, and language with the left (e.g., Tervaniemi et al., 2000; Zatorre, Evans, Meyer, & Gjedde, 1992), contemporary accounts posit that neural mechanisms are not

* Corresponding author at: Institute for Learning and Brain Sciences, Campus Box 357988, University of Washington, Seattle, WA 98195-7988, United States. Tel.: +1 206 221 6415; fax: +1 206 221 6472.
E-mail address: ginacc@u.washington.edu (G.C. Lebedeva).

as clearly defined as traditionally assumed, and that listening to both music and speech activates multiple and overlapping areas associated with multi-sensory perception (Altenmuller, 2001; Baeck, 2002; Kuhl & Damasio, in press). In approaching this topic, two main questions have been investigated in adults: (1) are the "tune" and "text" of a song stored and processed independently, or in an integrated manner (see Schön, Gordon, & Besson, 2005)? and (2) does the addition of melody enhance memory for words (Rainey & Larsen, 2002; Thaut, Peterson, & McIntosh, 2006; Yalch, 1991)? Studies attempting to answer the first question have produced evidence for both the independence and integration hypotheses, depending on the tasks involved, the measures used (i.e., what level of processing is examined), and the neurological status of the participants (Schön et al., 2005). Evidence is similarly equivocal regarding the second question. Given that neither has been systematically investigated in infants (Peretz & Coltheart, 2003; Trainor, 2005), we briefly review the literature on adults.

Before reviewing the more limited research on lyrics and melody in songs, it is important to note that the notion of perceptual integrality has been widely studied in many other domains. For instance, there are many variations of the well-established Garner interference effect, which demonstrates how the perception of multiple cues within the same signal may or may not influence each other, via facilitation, interference, or neither (Garner & Felfoldy, 1970). Even though features of a stimulus, such as pitch and loudness in sound, as well as hue, saturation, and brightness in vision, can be considered to be separate psychological dimensions, they are often unable to be fully processed completely independently (Kemler Nelson, 1993). Empirical evidence shows that changes in loudness for example can influence pitch judgments, and that there are asymmetrical effects on these perceptions based on whether loudness or pitch are increasing or decreasing in value (Neuhoff, Kramer, & Wayand, 2002). This body of literature has set the stage for investigating similar questions pertaining to features within combinations of speech and music.

1.1. Adult perception of lyrics and melody in songs

One line of research in adults has examined whether lyrics and melody are perceptually separable, when provided with multiple exemplars of each component. A pattern of findings across several studies has been collectively referred to as the integration effect, which purports that lyrics and melody of a song are perceived as a "mixed" mental representation that is not a mere concatenation of separate verbal and musical codes. For instance, some studies have found that adults demonstrate improved recognition of a given feature (lyrics or melody) when both the melody and lyrics of a song are coupled in the same combination in which they were originally learned (i.e., "congruent" pairings), as opposed to when melody and lyrics are mismatched (e.g., lyrics paired with a melody from another song), or when a feature is presented alone (Crowder, Serafine, & Repp, 1990; Hebert & Peretz, 2001; Samson & Zatorre, 1991; Serafine, Crowder, & Repp, 1984; Serafine, Davidson, Crowder, & Repp, 1986). Similarly, Bigand, Tillmann, Poulin, D'Adamo, and Madurell (2001) found that even when instructed to attend to either the lyric or melodic content, both trained musicians and non-musicians were faster at identifying the final vowel of a musical passage when it was sung on a harmonically congruous note, compared to a less congruous note, possibly reflecting integration between phonemic and harmonic processing. This basic finding was replicated in a lexical decision task involving more subtle harmonic violations (Poulin-Charronnat, Bigand, Madurell, & Peereman, 2005). Available evidence suggests that the integration effect is also present in preschool-aged children across cultures (Chen-Hafteck, 1999), but to a significantly lesser extent than in adults (Morrongiello & Roes, 1990). Related to the notion of perceptual integration is bidirectional association between familiar melodies and lyrics, where one feature heard in isolation activates components of the other in auditory priming tasks (Peretz, Radeau, & Arguin, 2004).

In the recognition and recall studies mentioned above, a perceptual asymmetry exists in which listeners consistently show evidence that processing lyrics, but not always melody, is obligatory at some level. For instance, Peretz et al. (2004) found that lyrics were easier to recognize than melodies, despite manipulations to decrease this asymmetry such as masking and distortions in temporal ordering. Similarly, both Crowder et al. (1990) and Serafine et al. (1984, 1986) demonstrated that listeners simply cannot "ignore" the lyrics despite explicit instructions to attend only to the melody, and that recognition performance is better in original melody-lyrics contexts compared to melodies with nonsense lyrics. Carrel, Smith, and Pisoni (1981) demonstrated that adults were slower at classifying syllables based on pitch when there was random vowel variation, but that classifying syllables based on vowels was not affected by pitch variation. In Morrongiello and Roes's study (1990), 4-year-old children were more likely to identify two songs as the "same" if at least the words were the same, and as "not at all the same" if the words were different, regardless of melody congruency. The asymmetrical patterns in these studies demonstrate that multiple cues can be integrated in perception, in that one dimension is difficult to explicitly separate from the other.

1.2. Evidence for "independence"

An alternative account proposes that discrete neural substrates are dedicated to musical and linguistic processing, reflecting an inherent biological foundation for music within the brain. The independent hypothesis is primarily supported by clinical data showing that people with neurological damage resulting in amusia (difficulty with music processing not accounted for by cognitive or linguistic deficits) can recognize spoken lyrics to a familiar song but cannot recognize the corresponding melodies of the songs when sung on "la" (Peretz, 1996; Peretz, Kolinsky, Tramo, & Labrecque, 1994). From these and similar clinical case studies evidencing dissociations between lyric and melodic perception (Hebert & Peretz, 2001; Samson & Zatorre, 1991), the modular processing model posits that the acoustic-to-phonological conversion from which semantic meaning is later derived, takes place separately but in parallel to pitch and temporal processing from which the

melodic meaning is derived (Peretz & Coltheart, 2003). Within speech processing, other types of brain damage, such as from a stroke, can result in selective impairment of segmental or suprasegmental features. For instance, some patients with primary progressive aphasia may show relatively few segmental errors, while making many errors in word stress placement (Janβen & Domahs, 2008), a condition that closely aligns with the distorted prosodic symptoms present in aprosodia (Ross & Monnot, 2008).

Support has also been gathered from non-impaired populations. Using electrophysiological measures. Besson, Faieta, Peretz, Bonnel, and Requin (1998) found that when listening to opera excerpts containing either an incongruous word or pitch on the final note of a phrase, professional musicians showed the classic N400 response for incongruous semantic content, and the P300 response for incongruous pitches. Because these components were identical to those elicited in speech-only or melody-only conditions, the authors concluded that semantic and harmonic content are processed independently of each other, even when both occur within the same signal. In addition, the size of neural responses to a double-violation was equal in amplitude to the sum of responses to the violations in isolation. In an analogous behavioral study, Bonnel, Faita, Peretz, and Besson (2001) found that there was no performance deficit when participants were asked to detect both semantic and harmonic values of the final note of an opera excerpt simultaneously, compared to individually. The lack of difference between the single- and dual-task conditions led to the conclusion that the semantic and melodic features are processed independently.

1.3. Multi-dimensional auditory perception infants

Given that typical stimuli in the infant's everyday experiences are inherently multi-dimensional, the ability to integrate multiple cues within a signal is an important milestone in development (Mattys, White, & Melhorn, 2005; Morgan & Saffran, 1995). In considering the perception of song components from a developmental standpoint, one assumption may be that infants start with melody and lyrics as separable, perhaps with one type of pattern being more salient than the other, then gain the ability to integrate, as their processing and memory capacity develops. However, it may be just as plausible to assume that infants actually start with an integral percept, and only later learn to separate lyrics from melody after sufficient experience with both music and language.

In light of these questions, most research to date on multi-dimensional acoustic perception in infants has focused on manipulating variables within speech or music, but not together. Previous work has demonstrated that infants can categorize complex auditory signals, speech and non-speech, based on single dimensions while ignoring irrelevant information on other dimensions, suggesting it is possible that two dimensions can be perceived independently of one another. For instance, by 6 months, infants can categorize vowels based on phonetic content despite talker variability (Kuhl, 1979; Kuhl, 1983). At 3 months of age, infants recognize vowel similarity despite pitch contour variability (Marean, Werner, & Kuhl, 1992). In another study, 7–9-month-olds categorized musical sequences based on rhythmic structure, despite random variations in melody (Trehub & Thorpe, 1989).

Other evidence that cues may be separable for infants comes from more recent studies that have investigated how infants infer word boundaries in running speech when there are multiple conflicting cues that are present. It is important to note that different types of suprasegmental cues, such as pitch and stress, are not necessarily equitable either within or across the domains of speech and music. However, the findings from the emerging body of literature in this section support the conclusion that suprasegmental and segmental properties within the same signal may be differentially attended to by infants. Specifically, three independent studies found evidence that 8–11-month-old infants find rhythmic cues, particularly stress patterns, in a speech stream more salient than the statistical structure of the segments within that stream (8 months: Johnson & Jusczyk, 2001; 9 months: Thiessen & Saffran, 2003; 11 months: Johnson & Seidl, 2009). Younger infants however, appear to rely relatively more on statistical cues than stress cues in segmentation (Thiessen & Saffran, 2003), though stress still provides a reliable means of segmentation (Thiessen & Saffran, 2007). These findings are consistent with the interpretation that through development, infants are better able to rely on a broader array of possible segmentation cues. In a more direct test of this hypothesis, Seidl (2006) found that 6-month-olds rely on a combination of the pitch and either pause duration or pre-boundary lengthening to successfully segment syntactic clauses from running speech. Thus, by halfway through their first year, infants have begun to show that not all cues are treated equivalently from a perceptual standpoint—some cues are weighted more strongly than others.

Other data from infants are consistent with the asymmetrical perception found in adults, where phonetic information is processed in an obligatorily manner when it is embedded anywhere in the signal. For example, infants 2–3 months old can attend to a phonetic change in vowels while ignoring irrelevant variation in pitch, though they do not detect a change in pitch while the vowel is varied, suggesting a primacy for phonetic versus pitch changes (Kuhl & Miller, 1982). These findings of perceptual asymmetry fit well with the notion that from early infancy, speech enjoys a "privileged" perceptual status, which is thought to be an adaptive feature pertinent to language acquisition (Kuhl & Miller, 1982; Vouloumanos & Werker, 2007).

1.4. Music as a mnemonic

Do songs facilitate memory for words? There is evidence of improved text recall in the context of songs compared to speech (Jellison, 1976; Rainey & Larsen, 2002; Wallace, 1994). One explanation of this facilitation effect is that multiple cues that are organized melodically, rhythmically, and linguistically, which each have the opportunity to be triggered in memory, provide

increased probability of accurate recall of the material. In addition, the tight temporal contingency between two or more events (as in lyrics and melody) increases the likelihood that they will become co-represented in memory (Crowder et al., 1990). While this facilitation effect is often exploited, such as through commercial jingles and early children's songs (Calvert, 2001; Yalch, 1991), as well as in clinical settings pertaining to cognitive and language rehabilitation (Brotons & Koger, 2000; Racette, Bard, & Peretz, 2006), constraints have been recognized. For example, both the melody and rhythm components must be either familiar (Gfeller, 1983) or sufficiently learned (Wallace, 1994) to produce the effect, and when presentation rate is equated across both speaking and singing conditions, the advantage for sung over spoken lyrics disappears (Kilgour, Jakobson, & Cuddy, 2000). Rainey and Larsen (2002) found that the facilitation effect is more robust for long-term, as opposed to short-term, recall. In addition, listeners with formal musical training show a stronger facilitation effect than those without such training (Kilgour et al., 2000), suggesting an element of "expertise-based" perceptual learning. Finally, verbal material in a song is most likely to be processed at the phonetic, rather than semantic level, as evidenced by children's and adults' increased verbatim recall of lyrics compared to spoken text, but increased comprehension of content in spoken compared to sung contexts (Calvert, 2001; Calvert & Billingsley, 1998).

1.5. The current study

The work reported here attempts to integrate the key patterns of results described above by posing two research questions in the context of infants' perception of lyrics and melody. Specifically, are infants able to perceptually separate phonetic from melodic patterns within songs, and if so, is one more perceptually salient than the other? Further, does singing facilitate recognition of phonetic content compared to speaking alone? Based on previous literature, there are several plausible and theoretically interesting outcomes. One is that phonetic changes are more easily attended to in the presence of pitch variation than pitch changes in the presence of phonetic variation, as reported previously in infants (Kuhl & Miller, 1982). Such results would demonstrate that processing changes in one dimension amidst variation in a second exhibits perceptual asymmetries that favor phonetic sensitivity. This selectivity would support the notion that biases exist early in life as a result of biological predispositions, exposure and experience, which direct infants' attention to signals that carry meaning—thus granting speech a special status in relation to non-linguistic signals (Vouloumanos & Werker, 2004). Alternatively, infants may find melody the most salient feature, and attend to it at a cost to the processing of phonetic detail (Trehub, 1990). While no direct test of this has been published, such an outcome would be consistent with findings that infants are particularly sensitive to vocal intonation patterns in speech due to their exposure to these cues in utero. Based on prenatal experience with pitch and rhythm cues in speech, newborns are able to reliably discriminate languages based on suprasegmental, but not segmental, cues alone (McMullen & Saffran, 2004). Along these lines, other reports show evidence that 8–9-month-old infants find rhythmic cues, particularly stress patterns, in a speech stream more salient when compared to the statistical structure of the segments within that stream (Johnson & Jusczyk, 2001; Thiessen & Saffran, 2003).

Finally, we hypothesized that the ability to discriminate phonetic detail within sung strings would be enhanced compared to spoken strings. Though no previous studies have been published in infants using the song context, support for this comes from studies reporting that higher and more variable pitch is helpful to infants in speech segmentation and discrimination tasks. For instance, Karzon (1985) found that 1–4-month-old infants were able to discriminate phonetically similar tri-syllabic utterances ('marana' from 'malana') only when the middle syllable contained exaggerated prosody (higher stress, pitch, and amplitude). Likewise, in a vowel categorization task, infants 6–7 months old performed better when exaggerated pitch contours (as found in IDS) were present (Trainor & Desjardins, 2002). Similarly, Thiessen, Hill, and Saffran (2005) found that 6–7-month-old infants were better able to distinguish words from a nonsense speech stream, in which the only cues to word boundaries were the transitional probabilities between syllables, when the stream was presented with the higher and more varied intonation contours characteristic of IDS, compared to ADS. While these studies examine only spoken (not sung) speech, it may be the case that this is a more general property of auditory processing. If so, then it is likely that singing, which contains similarly exaggerated pitch contours, may positively impact phonetic performance.

To test the outlined hypotheses, we framed this series of experiments around infants' relative sensitivity to changes in the sequential order of either syllables or pitches within auditory strings. This approach is ecologically valid, in that ordering of syllable (or word) units largely defines the linguistic or musical syntax of the string (e.g., the monosyllables in the sentence "You two are mine for good" can be rearranged in at least four meaningful orders; the four bars of the "Big Ben" chime arrangement have different configurations of the same four pitches, indicating various quarters of the hour). Two experiments were conducted. In both, infants were familiarized with a spoken or sung sequence, and their preference was subsequently tested for familiar versus novel (rearranged) sequences.

2. Experiment 1

In this experiment, 11-month-old infants were tested on pitch and phonetic changes presented in isolation. This age was chosen based on previous literature discussed above, suggesting that a more mature ability to attend to and process multiple cues within a signal is likely to be present. In addition, infants at this age are typically nearing production of their first spoken words, at which point perceptual weighting of phonetic versus pitch cues are particularly interesting. Each infant participated in two conditions in counterbalanced order: in the Melody Only condition, the task was to detect an order change in the note sequence of a four-note melody sung on the syllable /la/; in the Lyrics Only condition, the task was to detect a change

in syllable order in a spoken nonsense four-syllable string without pitch exaggeration. After the familiarization phase in each condition, the familiar stimulus and a novel stimulus (in which the ordering of the last three elements in the string was novel) were presented to the infant, and looking time to each type of stimulus was recorded. A significant preference for either the novel or familiar stimulus may be interpreted as evidence of discrimination. We predicted that infants would prefer novel strings for both melodic and phonetic changes, as simpler tasks and stronger encoding generally result in a novelty preference.

2.1. Methods

2.1.1. Participants

The sample consisted of 20 infants (11 girls) between the ages of 10.7 and 11.1 months ($M = 10.9$, SD = .10). Data from an additional six infants were excluded because they only completed the first of the two conditions ($n = 6$, 5 of whom had the speaking condition first) due to fussiness. According to parental report, all infants were delivered within 2 weeks of term, had no history of hearing problems, and were free of ear infections for 2 weeks prior to the time of testing. All were from monolingual English homes representative of the population, and after informed consent, parents or caregivers received compensation for participating.

2.1.2. Stimuli

Digital recordings were made in a sound attenuated booth, with 16-bit resolution and a sampling rate of 44 kHz. An adult native English speaker trained in phonetics and singing served as the voice for all the recorded stimuli. She produced numerous tokens of the songs and speech while listening to a pure tone version of the song (or a metronome during speaking), over headphones to facilitate tuning and timing of the recordings. After several recordings were made, the best were chosen based on accuracy of pitch, duration, and amplitude evenness, as measured aurally and by the examination of the spectrograms and waveforms. All syllables in both conditions were ∼400 ms long ($M = 405$ ms, SD = 11 ms), concatenated without pauses. A 600 ms pause was inserted between each syllable string, using Praat acoustic analysis software (Boersma & Weenink, 2007). Five naive adult listeners rated each token as at least "5" on a rating scale of naturalness, where 1 is 'very unnatural' and 7 is 'very natural.' All syllables were equated in amplitude, and were played at approximately 70 dB(A) SPL, measured at the approximate position of the infant's head.

The Melody Only familiarization stimulus consisted of a four-note melody sung on the syllable/la/. Four different recordings of this melody were concatenated and used as familiarization stimuli, and an additional recording served as the familiar test stimulus. Pitches for the familiarization melody were chosen pseudo-randomly from an E-Major scale:[1] A_{220}-$F\#_{185}$-E_{329}-B_{247} (subscript refers to the fundamental frequency, rounded to the nearest Hertz), creating a high-low-high-low melodic contour. The novel test stimulus was a re-ordering of the last three notes of the familiar stimulus, such that the contour was inverted: A_{220}-B_{247}-$F\#_{185}$-E_{329} (low-high-low-high). The first pitch remained the same to reduce potential reliance on the initial pitch to distinguish between stimuli. Contrasting melodic contours were used based on previous data suggesting that infants 9–10 months old more easily detect new melodies (6-note sequences, consisting of novel pitches) that contain contour violations compared to one that preserves contours (Trehub, Bull, & Thorpe, 1984; Trehub & Trainor, 1990).

The Lyrics Only familiarization stimulus was a four-syllable nonsense string,/gobiratu/. Four recordings were presented during familiarization, and one additional recording served as the familiar test stimulus. The novel test stimulus contained re-ordered syllables, with the first syllable constant:/goratubi/. Syllables were chosen to represent different place, manner, and voicing features in the consonants (without sibilants) and the point vowels, as well as a rounded dipthong. Acoustic values (voice onset time, formant frequencies and transitions) were comparable across like-tokens. Each syllable contained a slightly falling pitch contour starting at a mean frequency of 266 Hz (SD = 14.8) and ended at 216 Hz (SD = 2.6), with a mean frequency of 230 Hz (SD = 8.75), which is in between a typical infant-directed and adult-directed fundamental frequency.[1] A falling pitch contour was chosen to simulate more naturalistic speech compared to a steady monotone.

2.1.3. Procedure

The procedure was modeled after the head-turn preference paradigm (Nelson, Jusczyk, Mandel, Myers, Turk, & Gerken 1995; Thiessen et al., 2005), and used custom-made stimulus delivery and data collection software. Infants were individually tested in a sound attenuated booth while seated on a parent's or caregiver's lap. A hidden video camera recorded the infant. An experimenter in a room adjacent to the booth observed the infant on a video screen and performed online coding of the direction of head-turns. Both the experimenter and the parent were blind to the type of stimulus being delivered, listening to masking music over headphones. The order of condition presentation (Melody Only or Lyrics Only) was counterbalanced across participants.

The familiarization phase began with a blinking tube of white lighting, 30 cm long placed vertically on a panel 70 cm in front of the infant. Once the infant's gaze was directed forward at the light, the familiarization stimuli began playing

---

[1] In order to prevent inclusion of pitches that were unreasonably low or high, this scale was chosen based on data collected from 5-min naturalistic infant- and adult-directed speech from 30 mothers. The mean fundamental frequency of the mothers' point vowels (/i/,/a/,/u/) was 213 Hz ($M_{ID} = 249$ Hz; $M_{AD} = 176$ Hz), which corresponded approximately to the note $A_{220\,Hz}$; this is the middle (4th note) of the E-Major scale.
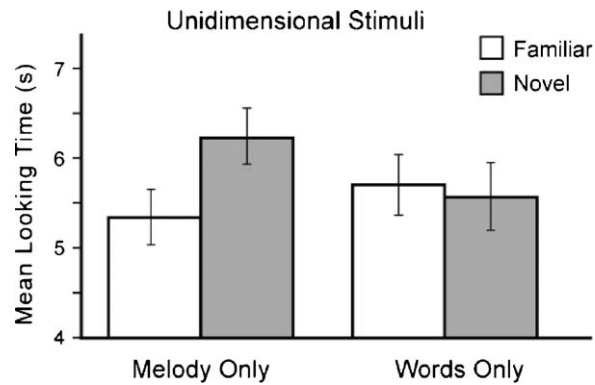
Fig. 1. Mean looking times for experiment 1 (+SE) for familiar and novel stimuli in Melody Only or words only (uni-dimensional) conditions.

simultaneously from two hidden speakers placed at 90° to the right and left of the infant. The sounds played continuously for 1 min, while the center light continued blinking to encourage attentiveness. However, the sound and light stimuli were constant, and non-contingent on infant behavior or response.

Immediately after the familiarization phase, the sounds stopped, and the center light blinked until the infant looked forward; this signal started a trial. When the infant's gaze was at center, the experimenter pressed a key and a randomly selected side light began to blink (and the center light stopped blinking). When the infant made a head-turn of at least 30° toward the blinking light, the experimenter pressed another key, and one of the test stimuli was played from the speaker behind the blinking light. The sound continued repeating until either 15 consecutive seconds of looking time had occurred, or until the infant looked away, at which point the experimenter released the key, which stopped the sound.[2] The center light again began blinking, indicating the start of another trial. Twelve test trials were then administered: six were presentations of the familiar stimulus (Melody Only or Lyrics Only), and six were the novel stimulus (with the order of the last three units rearranged). Trials were presented in pseudo-random order, with six trials (three familiar and three novel) presented from each side speaker. After 12 trials were complete, the familiarization phase immediately began for the other condition, followed by its testing phase. The dependent measures used in analyses were the amount of looking times toward the familiar and novel stimuli.

2.2. Results and discussion

In the Melody Only condition, the mean looking times (and standard errors) for the familiar and novel stimuli were 5.34 s (.30) and 6.23 s (.32), respectively; means for the Lyrics Only condition were 5.69 s (.34) and 5.57 s (.37), as depicted in Fig. 1. A repeated-measures ANOVA was conducted using condition and stimulus type as within-subjects factors. There was no main effect, but the interaction was significant, $F(1,19) = 4.29$, $p = .05$, $\eta_p^2 = .18$, driven by the infants' preference for the novel melody in the Melody Only condition and their lack of preference in the Lyrics Only condition. This pattern of results suggests that infants can readily discriminate between two 4-note melodies sung on "la" that differ in the sequential order of the last three pitches, but do not discriminate between two spoken strings that contain a change in syllable order, at least when each syllable is spoken with the same slightly falling intonation contour.

The lack of preference between the two 4-syllable strings is in accordance with data from 6-month-old infants who showed difficulty in processing sequential information in tri-syllables (Morgan & Saffran, 1995). Though we predicted that infants would demonstrate a preference based on previous evidence of infants using transitional probabilities to recognize units in streams of both syllables (Saffran, Aslin, & Newport, 1996) and tones (Saffran, Johnson, Aslin, & Newport, 1999), it is possible that because our stimuli were presented as strings with pauses, infants were not relying on statistical properties to segment the sounds, and were instead attending to more global characteristics of each unit (Trehub, 1990). These and other possible interpretations will be discussed further in Section 4.

Given the apparent perceptual asymmetry where detecting a change in pitch order was easier than detecting a change in syllable order, what would happen when both dimensions are present within a signal? To test this, experiment 2 investigated whether infants at this age would continue to show the same asymmetrical melody-favoring pattern when hearing songs with words, or whether they would effectively ignore the melody and be relatively more sensitive to changes in the lyrics instead.

---

[2] Typically, studies using this paradigm have included a 2-s criterion for looking away before stopping the stimulus. Pilot experimentation with this criterion led to a disproportionate number of infants attending to one or both of the speakers for <3 s total, possibly because our stimuli consisted of a 1.6-s long repeated string with a 0.6-s pause in between, and not a continuous stream as in previous studies.

3. Experiment 2

Experiment 2 utilized the identical melody and syllable order changes from the previous experiment, but in the context of song stimuli, in which both melody and lyrics were present in the same signal. All infants, 11 months old, were familiarized with the same song, and that song served as the "familiar" stimulus for all groups. For the "novel" stimulus in the test phase, one group of infants heard the familiar melody with re-ordered syllables ("New Lyrics"), and the other group heard the familiar syllable sequence with re-ordered pitches ("New Melody"). As in the previous experiment, the first note and syllable of the songs remained constant to prevent changes in the initial unit from being a cue to the status of the stimulus. Based on the hypotheses outlined above, we predicted that one of two alternate outcomes were plausible. If the melodic pattern is more perceptually salient and more easily attended to, infants would continue to show a preference when presented with a new melody, regardless of whether the syllables remain constant (experiment 1) or variegated (experiment 2). Alternatively, it may be the case that the pitch cues in melody actually enhance phonetic discrimination performance (which could plausibly be due to increased attention that pitch variation and singing tend to command, or to acoustic-phonetic differences in sung versus spoken speech), similar to recent findings on the facilitative effects of infant-directed speech, as described above. In this case, we would expect the infants to demonstrate a preference when presented with new lyrics, despite a melody that is held constant. This outcome pattern would suggest that phonetic information within a melody cannot be "ignored" even when a perceptually salient pitch change is introduced within the same signal.

3.1. Methods

3.1.1. Participants

Forty infants (11 girls) between the ages of 10.6 and 11.2 months ($M = 10.85$, SD = .14) participated in experiment 2. Data from an additional 10 infants were excluded due to fussiness ($n = 8$, 6 of whom were assigned to the New Lyrics condition), equipment failure ($n = 1$), and for showing a preference direction more than 3 standard deviations from the mean ($n = 1$ in the New Melody condition). All other participant criteria were identical to those in experiment 1.

3.1.2. Stimuli

Stimuli were recorded, prepared and presented in the manner as described in the previous experiments, with the exception that the woman (a trained vocalist) sung the melodies with the variegated syllables as full songs. The familiarization song consisted of the familiarization melody sung with the familiarization lyrics from experiment 1. The novel songs consisted of either the re-ordered pitches or syllables of the previous experiments, while the other dimension remained constant. Thus, infants heard identical melody and syllable changes as in the previously tested uni-dimensional contexts. Four tokens served as the familiarization stimuli, and the familiar stimulus in the test phase was a novel token of the familiar song. Five naive adult listeners correctly identified all syllables in all song stimuli, indicating that the phonetic content was intelligible.

3.1.3. Procedure

The procedure was identical to that used in experiment 1, with the exception that each infant participated in only one condition. Infants were randomly assigned to either the New Melody condition, in which the novel stimulus consisted of a re-ordering of the final three pitches of the stimulus, or the New Lyrics condition, in which the novel stimulus contained a re-ordering of the last three syllables. Thus, all infants were familiarized with the same song during the 1-min familiarization phase, while the test phase consisted of either a New Melody or New Lyrics.

3.2. Results and discussion

Results are summarized in Fig. 2. In the New Melody condition, infants looked at the familiar song for 6.66 s (SE = .43) and at the novel song for 6.37 s (SE = 0.43), with 11 of the 20 infants showing a novelty preference (55%). This difference was not significant based on a two-tailed paired t-test, $t(19) = .66$, $p = .52$. Thus infants did not show evidence that they discriminated the songs based on their differing melodic contour, despite their ability to discriminate the identical change in melody when it was presented without variegated phonetic content in experiment 1.

In the New Lyrics condition, infants listened to the familiar song for 5.45 s (SE = .30) and to the novel song for 6.28 s (SE = .28). This difference approached significance in a two-tailed t-test, $t(19) = 1.77$, $p = .09$. Fifteen of 20 (75%) showed a novelty preference, which was statistically significant based on the binomial test, $p = .04$. Reexamination of the data revealed that one infant's mean looking time difference to the familiar and novel stimuli was 2.18 standard deviations away from the mean, with the next closest individual only 1.48 away; visual inspection of individual performance plots suggested that there were no such outliers in any of the other conditions across both experiments. A two-tailed t-test excluding this outlier showed that infants preferred listening to the song containing the novel (re-ordered) over the familiar lyrics, $t(18) = 2.52$, $p = .02$, $d = .49$; the binomial test remained significant, $p = .02$. To confirm that this subject was indeed acting as an outlier, both conditions in experiment 2, as well as both conditions in experiment 1, were re-examined by removing each subject, one at a time, from the described comparison analyses. In every single case except for the aforementioned participant, the patterns of significance did not change. The remaining analyses exclude data from this infant in both conditions; resultant mean looking times are depicted in Fig. 2.
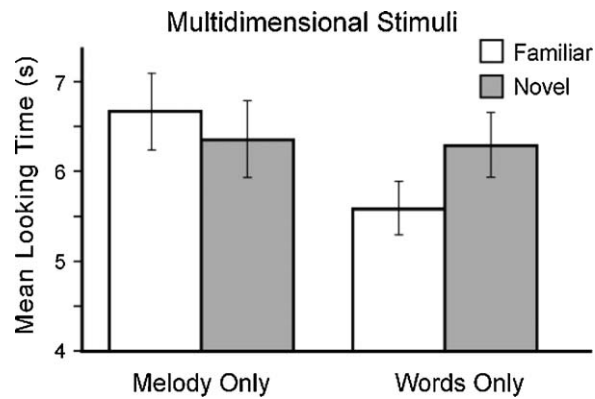
Fig. 2. Mean looking times for experiment 2 (+SE) for familiar and novel stimuli in new melody or new lyrics (multi-dimensional) conditions.

To test whether performance significantly differed between the New Melody and the New Lyrics conditions in experiment 2, we conducted a 2×2 (condition $\times$ item type) repeated-measures ANOVA, and found no main effects but a marginally significant interaction, $F(1, 37) = 3.29$, $p = .07$, $\eta_p^2 = .08$, again reflecting the stronger novelty preference in the New Lyrics condition, compared to the lack of any preference in the New Melody condition. Because the infants showed a novelty preference in the Melody Only condition of experiment 1 but did not show any preference for either melody in the New Melody condition of experiment 2, we conducted a 2×2 (condition $\times$ item type) repeated-measures ANOVA, and found a significant interaction, $F(1, 38) = 4.81$, $p = .03$, $\eta_p^2 = .11$. This result suggests that even though the infants could discriminate two melodies sung with the redundant syllable "la," they did not show evidence of discrimination when the identical melody change was presented with varied syllables.

Finally, to test the hypothesis that adding melody to phonetic strings can facilitate phonetic discrimination compared to speaking, a 2×2 (condition $\times$ item type) mixed ANOVA was conducted on the data from the Lyrics Only condition from experiment 1 and the New Lyrics condition from experiment 2. These conditions are easily comparable because each contains identical phonetic re-ordering, but they are spoken in experiment 1 and sung to the same melody in experiment 2. There was no main effect of item type, but the interaction between item and condition approached significance, $F(1, 37) = 2.93$, $p = .09$, $\eta_p^2 = .07$. While this result has a 9% likelihood of being a product of chance, this marginal effect demonstrates the possibility that the melodic context facilitated recognition of the phonetic content.

4. General discussion

This study examined infants' ability to detect phonetic and melodic changes within spoken and sung strings at the age of 11 months. Literature on adults has debated whether lyrics and melodies of songs are perceived as an integrated unit or as two separate sources of information, with one component more perceptually salient than the other. This study examined the adult predictions in studies of infants. Second, it has been suggested that melody can serve to enhance recall of words, and this study investigated whether phonetic details within sung syllables are easier for infants to recognize compared to the same syllables when they are spoken.

In experiment 1, infants robustly preferred the novel melody when represented by a change in pitch order of 4-note melodies sung on "la;" but they did not show evidence of discriminating a change in syllable order in two 4-syllable strings spoken with falling intonation. This finding suggests that detecting a pitch change with redundant syllables is a relatively easier task compared to detecting phonetic changes in syllable strings. Experiment 2 presented infants with multi-dimensional stimuli of melodies with variegated syllables, and tested infants' detection of a change in one dimension (either syllable or pitch order). Here, two important patterns emerged. First, infants no longer detected a change in melody when variegated syllables were present. This is in contrast to their ability to detect the identical melody change with redundant syllables in experiment 1. Thus, the data suggest that infants' attention seemed to be drawn to the phonetic content—the lyrics—at some level, as opposed to simply ignoring the phonetic content and discriminating the stimuli based on the pitch differences alone.

A second pattern emerged from the data. Our results suggest that phonetic redundancy facilitates the processing of a melodic line, a pattern consistent with other findings in speech (Kuhl & Miller, 1982). Infants showed more reliable detection of a change in syllable order within sung stimuli, but did not detect the identical phonetic changes in a spoken string in experiment 1. Taken together, these results are the first to support the hypotheses that (1) at 11 months, infants appear unable to "ignore" phonetic content in the context of perceptually salient pitch information and (2) that recognition of segmental features is improved in the context of sung, compared to spoken, presentations that do not contain varied pitch contours.

There are several patterns of results that warrant discussion. In the Lyrics Only condition in experiment 1, infants did not show evidence of discrimination between two 4-syllable strings that differ in the order of the last three syllables. This

seems initially surprising, given that even younger infants are sensitive to fine acoustic differences in speech (Jusczyk, 1999). One possible explanation is that this task required not just phonetic discrimination, but detecting sequential changes of syllables, which more heavily depends on higher cognitive skills of attending to, retaining, and recalling serial information. The ability to combine these develops over the first year; processing sequential information in tri-syllables, for example, has been shown to be a later-emerging skill (Morgan & Saffran, 1995). Another possible explanation for the lack of preference in the spoken condition may be that the spoken strings are less "natural" compared to the sung stimuli, due to the controlled pitch and timing characteristics of the speech. Because each syllable essentially had a sentential falling contour, infants (who prefer natural infant-directed speech) may have become bored hearing the same intonation pattern, and consequently showed no learning. While the contour controls were necessary to eliminate the possibility that an advantage for sung over spoken stimuli is due to temporal cues (Kilgour et al., 2000), finding the same pattern of results using more natural spoken stimuli (in both infant- and adult-directed registers, using an overall falling contour) would strengthen the conclusions of the present study, as well as other reports (Thiessen et al., 2005) regarding the facilitative effects of exaggerated pitch information on phonetic perception. Finally, another possible explanation for the lack of preference in the Lyrics Only condition is that the familiarization time was not long enough for the infants to fully learn the pattern. Results of our pilot testing had suggested that a longer familiarization time often resulted in a higher rate of attrition. Though a 1-min familiarization period is relatively short, other studies have demonstrated that infants can reliably detect phonological regularities in infant-directed speech after that amount of exposure (Thiessen et al., 2005).

Even though infants encoded a change in melodic information when it was presented in isolation, pairing the melodic information with syllables failed to produce evidence of encoding. This observed difference between experiment 1 and 2 could have been due to the increased complexity of the stimuli; holding the syllable order constant in experiment 1 made it easier for infants to focus on the melody, whereas the varied syllables in experiment 2 made the same task more difficult due to having more information to process. The finding provides some support for the notion that speech has a "privileged" perceptual status from early in life (Vouloumanos & Werker, 2007). The result suggests that in the context of song, which is designed to attract and maintain attention while communicating affect through melodic and rhythmic components, phonetic content that is present will also be processed.

4.1. The facilitation of phonetic recognition with exaggerated pitch cues

Because infants responded to the syllable order changes in the context of the sung but not in the spoken condition, we may infer that the addition of pitch cues enhanced infants' ability to detect changes in syllable order. This can be interpreted as evidence that when listening to songs, infants may attend to phonemes more than melody; when the lyrics are changed, recognition is enhanced relative to a melody change. These interpretations are consistent with reports that recall performance of song lyrics is comparable for both real words and nonsense words (Serafine et al., 1984, 1986), suggesting that phonetic, not semantic, features contribute to the facilitation effect. Similarly for preschool children, adding melody to text increases the verbatim recall of the text, but not necessarily the semantic content, compared to prose presentation (Calvert & Billingsley, 1998; Yalch, 1991). In other words, young children tend to be able to repeat the words of a song, without necessarily being able to rely the meaning of the lyrics. The melody might facilitate a more "superficial" level of encoding at the phonological level, but not the lexeme itself (see Brotons & Koger, 2000; Racette et al., 2006 for similar effects in people with dementia and aphasia, respectively).

Our data supporting the facilitation effect of added melody is in line with other behavioral findings from infants, which suggest that exaggerated pitch information within a signal can enhance language-learning skills, such as word segmentation based on transitional probability tracking (Thiessen et al., 2005) and vowel categorization (Trainor & Desjardins, 2002). This facilitation is likely to be based primarily on pitch exaggeration, as opposed to temporal and amplitude emphasis (Fernald & Kuhl, 1987). Recently Bergeson and Trehub (2007) have suggested that mothers' tendency to use distinctive "speech tunes" in interactions with their infants may enhance infants' recognition of words or phrases after a period of exposure. From this, a working hypothesis is that performance in detecting changes in syllable order would be improved in both the context of infant-directed speech and singing compared to adult-directed speech, even with controlled temporal cues. This result would further emphasize the important role of pitch variation in phonetic learning.

The observed facilitation effect of the melody may be due to the pitch contour's role in attracting, maintaining, or enhancing overall attention, which then facilitates recognition of phonetic patterns (Fernald, 1989; Fernald, 1992). Similarly, the presence of a melodic line within a signal may maximize infants' arousal (Shenfield, Trehub, & Nakata, 2003) and attention (Trainor, Clark, Huntley, & Adams, 1997), allowing for enhanced learning of concomitant input. An alternative explanation may be that the added pitch component provides an additional acoustic cue as part of the phonemes themselves, which helps differentiate the syllable strings in memory—the present studies cannot differentiate between these possible explanations. We also note that the current set of studies did not test naturalistic infant- or adult-directed speech, but instead maintained identical falling pitches on each syllable, using a median pitch that was typical of infant-directed speech. It has been suggested that the discrete, steady pitches present in singing, though they vary in height, could draw infants' attention away from the phonetic patterns contained within them (Trehub & Trainor, 1990); however, our data suggest that the addition of melody does not hinder phonetic recognition, but actually facilitates it.

4.2. Implications for developmental differences in pitch versus phonetic pattern recognition

It would be of value to investigate these findings across different developmental stages. Our own preliminary testing of the identical paradigm with 8-month-old infants showed no evidence of discrimination between either the melodies or the spoken strings, under the identical familiarization and testing conditions. This null result may be due to the relatively short familiarization time (1 min), which may not have been long enough for 8-month-old infants to adequately encode the pattern (Hunter & Ames, 1988). Alternatively, it could also be suggestive of important developmental changes regarding how pitch and phonetic patterns are attended to and encoded with increasing age. Language experience may produce developmental changes in the perceptual weightings of particular dimensions of sound in multi-dimensional stimuli, and this could in turn play an important role in learning (Johnson & Jusczyk, 2001; Thiessen & Saffran, 2003). Linguistic experience over time may produce increased attention to phonetic compared to pitch cues within a multi-dimensional (sung) signal. As infants approach the typical age for first words, they may be increasingly able to differentially attend to phonetic and pitch information, and this would support infants' attention toward phonetic cues that relate to lexical information (Johnson & Jusczyk, 2001; Thiessen & Saffran, 2003).

4.3. Summary

Classic issues in the adult literature pertaining to how lyrics and melodies of songs are perceived led us to approach the developmental question of how infants process multiple cues present within speech and song. Based on our data, we offer two specific findings that contribute to the literature on the development of auditory perception and how it relates to language acquisition. First, we demonstrate that, for 11-month-old infants, phonetic content within a signal is likely to take perceptual precedence over pitch content, at least when both are present within a song: infants at this age are not able to "ignore" the lyrics even when embedded in a melody that is easily discriminable in isolation. Second, our findings suggest that exaggerated pitch patterns in songs may facilitate recognition of the phonetic content within them when compared to a speech context that lacks salient pitch information. These findings highlight the importance of considering stimulus complexity when examining infants' perceptual abilities, and also mirror the growing literature supporting the benefits of infant-directed speech in child language development.

Acknowledgements

References

Altenmuller, E. O. (2001). In R. J. Zatorre, & I. Peretz (Eds.), *How many music centers are in the brain?*.
Baeck, E. (2002). The neural networks of music. *European Journal of Neurology*, 9(5), 449–456.
Bergeson, T. A., & Trehub, S. E. (2007). Signature tunes in mothers' speech to infants. *Infant Behavior and Development*, 30(4), 648–654.
Besson, M., Faieta, F., Peretz, I., Bonnel, A., & Requin, J. (1998). Singing in the brain: Independence of lyrics and tunes. *Psychological Science*, 9(6), 494–498.
Bigand, E., Tillmann, B., Poulin, B., D'Adamo, D., & Madurell, F. (2001). The effect of harmonic context on phoneme monitoring in vocal music. *Cognition*, 81(1), B11–B20.
Boersma, P., & Weenink, D. (2007). Praat: Doing phonetics by computer (version 4.3.26) [computer program]. Available from http://www.praat.org.
Bonnel, A.-M., Faita, F., Peretz, I., & Besson, M. (2001). Divided attention between lyrics and tunes of operatic songs: Evidence for independent processing. *Perception and Psychophysics*, 63(7), 1201–1213.
Brotons, M., & Koger, S. M. (2000). The impact of music therapy on language functioning in dementia. *Journal of Music Therapy*, 37(3), 183–195.
Calvert, S. L. (2001). Impact of televised songs on children's and young adults' memory of educational content. *Media Psychology*, 3(4), 325–342.
Calvert, S. L., & Billingsley, R. L. (1998). Young children's recitation and comprehension of information presented by songs. *Journal of Applied Developmental Psychology*, 19(1), 97–108.
Carrel, T. D., Smith, L. B., & Pisoni, D. B. (1981). Some perceptual dependencies in speeded classification of vowel color and pitch. *Perception and Psychophysics*, 29(1), 1–10.
Chen-Hafteck, L. (1997). Music and language development in early childhood: Integrating past research in the two domains. *Early Child Development and Care*, 130(1), 85–97.
Chen-Hafteck, L. (1999). Discussing text-melody relationship in children's song-learning and singing: A Cantonese-speaking perspective. *Psychology of Music*, 27(1), 55–70.
Crowder, R. G., Serafine, M. L., & Repp, B. H. (1990). Physical interaction and association by contiguity in memory for the words and melodies of songs. *Memory and Cognition*, 18(5), 469–476.
Fernald, A. (1989). Intonation and communicative intent in mothers' speech to infants: Is the melody the message? *Child Development*, 60(6), 1497–1510.
Fernald, A. (1992). Meaningful melodies in mothers' speech to infants. In H. Papousek, U. Jurgens, & M. Papousek (Eds.), *Nonverbal vocal communication: Comparative and developmental approaches (pp. xvi, 303)*. Cambridge: Cambridge University Press.
Fernald, A., & Kuhl, P. K. (1987). Acoustic determinants of infant preference for mothers speech. *Infant Behavior and Development*, 10(3), 279–293.
Garner, W. R., & Felfoldy, G. L. (1970). Integrality of stimulus dimensions in various types of information processing. *Cognitive Psychology*, 1(3), 225–241.

Gerken, L., & Aslin, R. N. (2005). Thirty years of research on infant speech perception: The legacy of Peter W. Jusczyk. *Language Learning and Development*, *1*(1), 5–12.

Gfeller, K. (1983). Music mnemonics as an aid to retention with normal and learning disabled students. *Journal of Music Therapy*, *20*(4), 179–189.

Hebert, S., & Peretz, I. (2001). Are text and tune of familiar songs separable by brain damage? *Brain and Cognition*, *46*(1–2), 169–175.

Hunter, M. A., & Ames, E. W. (1988). A multifactor model of infant preferences for novel and familiar stimuli. In C. Rovee-Collier, L. P. Lipsitt, & H. Hayne (Eds.), *Advances in Infancy Research* (pp. 69–95). Norwood, NJ: Ablex Publishing.

Janβen, U., & Domahs, F. (2008). Going on with optimised feet: Evidence for the interaction between segmental and metrical structure in phonological encoding from a case of primary progressive aphasia. *Aphasiology*, *22*(11), 1157–1175.

Jellison, J. A. (1976). Accuracy of temporal order recall for verbal and song digit-spans presented to right and left ears. *Journal of Music Therapy*, *13*(3), 114–129.

Johnson, E. K., & Jusczyk, P. W. (2001). Word segmentation by 8-month-olds: When speech cues count more than statistics. *Journal of Memory and Language*, *44*(4), 548–567.

Johnson, E. K., & Seidl, A. H. (2009). At 11 months, prosody still outranks statistics. *Developmental Science*, *12*(1), 131–141.

Jusczyk, P. W. (1999). How infants begin to extract words from speech. *Trends in Cognitive Sciences*, *3*(9), 323–328.

Karzon, R. G. (1985). Discrimination of polysyllabic sequences by one- to four-month old infants. *Journal of Experimental Child Psychology*, *39*(2), 326–342.

Kemler Nelson, D. G. (1993). Processing integral dimensions: The whole review. *Journal of Experimental Psychology: Human Perception and Performance*, *19*(5), 1105–1113.

Kilgour, A. R., Jakobson, L. S., & Cuddy, L. L. (2000). Music training and rate of presentation as mediators of text and song recall. *Memory and Cognition*, *28*(5), 700–710.

Kuhl, P. K. (1979). Speech perception in early infancy: Perceptual constancy for spectrally dissimilar vowel categories. *Journal of the Acoustical Society of America*, *66*(6), 1668–1679.

Kuhl, P. K. (1983). Perception of auditory equivalence classes for speech in early infancy. *Infant Behavior and Development*, *6*(3), 263–285.

Kuhl, P. K., & Damasio, A. (in press). Language. In E. R. Kandel, J. H. Schwartz, T. M. Jessell, S. Siegelbaum, & J. Hudspeth (Eds.), *Principles of neural science*. New York: McGraw Hill.

Kuhl, P. K., & Miller, J. D. (1982). Discrimination of auditory target dimensions in the presence or absence of variation in a second dimension by infants. *Perception and Psychophysics*, *31*(3), 279–292.

Marean, G., Werner, L. A., & Kuhl, P. K. (1992). Vowel categorization by very young infants. *Developmental Psychology*, *28*(3), 396–405.

Mattys, S. L., White, L., & Melhorn, J. F. (2005). Integration of multiple speech segmentation cues: A hierarchical framework. *Journal of Experimental Psychology: General*, *134*(4), 477–500.

McMullen, E., & Saffran, J. R. (2004). Music and language: A developmental comparison. *Music Perception*, *21*(3), 289–311.

Morgan, J. L., & Saffran, J. R. (1995). Emerging integration of sequential and suprasegmental information in preverbal speech segmentation. *Child Development*, *66*(4), 911–936.

Morrongiello, B. A., & Roes, C. L. (1990). Children's memory for new songs: Integration or independent storage of words and tunes? *Journal of Experimental Child Psychology*, *50*(1), 25–38.

Nelson, D. G., Jusczyk, P. W., Mandel, D. R., Myers, J., Turk, A., & Gerken, L. (1995). The head-turn preference procedure for testing auditory perception. *Infant Behavior and Development*, *18*(1), 111–116.

Neuhoff, J. G., Kramer, G., & Wayand, J. (2002). Pitch and loudness interact in auditory displays: Can the data get lost in the map? *Journal of Experimental Psychology: Applied*, *8*(1), 17–25.

Patel, A. D. (2003). In G. Avanzini, C. Faienza, D. Minciacchi, L. Lopez, & M. Majno (Eds.), *Rhythm in language and music. Parallels and differences*

Peretz, I. (1996). Can we lose memory for music? A case of music agnosia in a nonmusician. *Journal of Cognitive Neuroscience*, *8*(6), 481–496.

Peretz, I., & Coltheart, M. (2003). Modularity of music processing. *Nature Neuroscience*, *6*(7), 688–691.

Peretz, I., Kolinsky, R., Tramo, M., Labrecque, R., et al. (1994). Functional dissociations following bilateral lesions of auditory cortex. *Brain: A Journal of Neurology*, *117*(6), 1283–1301.

Peretz, I., Radeau, M., & Arguin, M. (2004). Two-way interactions between music and language: Evidence from priming recogntion of tune and lyrics in familiar songs. *Memory and Cognition*, *32*(1), 142–152.

Poulin-Charronnat, B., Bigand, E., Madurell, F., & Peereman, R. (2005). Musical structure modulates semantic priming in vocal music. *Cognition*, *94*(3), B67–B78.

Racette, A., Bard, C., & Peretz, I. (2006). Making non-fluent aphasics speak: Sing along! *Brain*, *129*(Pt 10), 2571–2584.

Rainey, D. W., & Larsen, J. D. (2002). The effects of familiar melodies on initial learning and long-term memory for unconnected text. *Music Perception*, *20*(2), 173–186.

Ross, E. D., & Monnot, M. (2008). Neurology of affective prosody and its functional-anatomic organization in right hemisphere. *Brain and Language*, *104*(1), 51–74.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926–1928.

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, *70*(1), 27–52.

Samson, S., & Zatorre, R. J. (1991). Recognition memory for text and melody of songs after unilateral temporal lobe lesion: Evidence for dual encoding. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*(4), 793–804.

Schön, D., Gordon, R. L., & Besson, M. (2005). Musical and linguistic processing in song perception. In G. Avanzini, L. Lopez, S. Koelsch, & M. Manjno (Eds.), *The neurosciences and music II: From perception to performance* (pp. 71–81). Annals of the New York Academy of Sciences.

Schön, D., Magne, C., & Besson, M. (2004). The music of speech: Music training facilitates pitch processing in both music and language. *Psychophysiology*, *41*(3), 341–349.

Seidl, A. H. (2006). Infants' use and weighting of prosodic cues in clause segmentation. *Journal of Memory and Language*, *57*(1), 24–48.

Serafine, M. L., Crowder, R. G., & Repp, B. H. (1984). Integration of melody and text in memory for songs. *Cognition*, *16*(3), 285–303.

Serafine, M. L., Davidson, J., Crowder, R. G., & Repp, B. H. (1986). On the nature of melody-text integration in memory for songs. *Journal of Memory and Language*, *25*(2), 123–135.

Shenfield, T., Trehub, S. E., & Nakata, T. (2003). Maternal singing modulates infant arousal. *Psychology of Music*, *31*(4), 365–375.

Tervaniemi, M., Medvedev, S., Alho, K., Pakhomov, S., Roudas, M., van Zuijen, T., et al. (2000). Lateralized automatic auditory processing of phonetic versus musical information: A PET study. *Human Brain Mapping*, *10*(2), 74–79.

Thaut, M. H., Peterson, D. A., & McIntosh, G. C. (2006). Temporal entrainment of cognitive functions: Musical mnemonics induce brain plasticity and oscillatory synchrony in neural networks underlying memory. In G. Avanzini, L. Lopez, S. Koelsch, & M. Manjno (Eds.), *The neuosciences and music II: From perception to performance* (pp. 243–254). New York, NY: New York Academy of Sciences.

Thiessen, E. D., Hill, E. A., & Saffran, J. R. (2005). Infant-directed speech facilitates word segmentation. *Infancy*, *7*(1), 53–71.

Thiessen, E. D., & Saffran, J. R. (2003). When cues collide: Use of stress and statistical cues to word boundaries by 7- to 9-month-old infants. *Developmental Psychology*, *39*(4), 706–716.

Thiessen, E. D., & Saffran, J. R. (2007). Learning to learn: Infants' acquisition of stress-based strategies for word segmentation. *Language Learning and Development*, *3*(1), 73–100.

Trainor, L. J. (2005). Are there critical periods for musical development? *Developmental Psychobiology*, *46*(3), 262–278.

Trainor, L. J., Clark, E. D., Huntley, A., & Adams, B. A. (1997). The acoustic basis of preferences for infant-directed singing. *Infant Behavior and Development*, *20*(3), 383–396.

Trainor, L. J., & Desjardins, R. N. (2002). Pitch characteristics of infant-directed speech affect infants' ability to discriminate vowels. *Psychonomic Bulletin and Review*, *9*(2), 335–340.

Trehub, S. E. (1990). The perception of musical patterns by human infants: The provision of similar patterns by their parents. In M. A. Berkley, & W. C. Stebbins (Eds.), *Comparative perception (Vol. 1, Basic Mechanisms, pp. 429–459)*. New York: John Wiley & Sons.

Trehub, S. E., Bull, D., & Thorpe, L. A. (1984). Infants' perception of melodies: The role of melodic contour. *Child Development*, *55*(3), 821–830.

Trehub, S. E., & Thorpe, L. A. (1989). Infants' perception of rhythm: Categorization of auditory sequences by temporal structure. *Canadian Journal of Psychology*, *43*(2), 217–229.

Trehub, S. E., & Trainor, L. J. (1990). Rules for listening in infancy. In J. T. Enns (Ed.), *The development of attention: Research and theory* (pp. 87–119). Amsterdam: Elsevier Science Publishers.

Vouloumanos, A., & Werker, J. F. (2004). Tuned to the signal: The privileged status of speech for young infants. *Developmental Science*, *7*(3), 270–276.

Vouloumanos, A., & Werker, J. F. (2007). Listening to language at birth: Evidence for a bias for speech in neonates. *Developmental Science*, *10*(2), 159–164.

Wallace, W. T. (1994). Memory for music: Effect of melody on recall of text. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(6), 1471–1485.

Yalch, R. F. (1991). Memory in a jingle jungle: Music as a mnemonic device in communicating advertising slogans. *Journal of Applied Psychology*, *76*(2), 268–275.

Zatorre, R. J., Evans, A. C., Meyer, E., & Gjedde, A. (1992). Lateralization of phonetic and pitch discrimination in speech processing. *Science*, *256*(5058), 846–849.