Infant Behavior & Development 33 (2010) 419–430

Contents lists available at ScienceDirect

Infant Behavior and Development

# 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*

ARTICLE INFO

*Article history:*
Received 27 November 2007
Received in revised form 8 October 2009
Accepted 12 April 2010

*Keywords:*
Melody
Lyrics
Infant-directed speech
Speech and song perception
Language acquisition
Infant music perception

ABSTRACT

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.



## 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).

0163-6383/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.infbeh.2010.04.006

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