Expressive timing and dynamics in infant-directed and non-infant-directed singing

TAKAYUKI NAKATA
Future University-Hakodate

SANDRA E. TREHUB
University of Toronto & Université de Montréal

Abstract—Expressive variations in timing and dynamics were examined in infant-directed performances of Twinkle, Twinkle, Little Star by mothers and in non-infant-directed performances by non-mothers. Mothers sang to their infants, and non-mothers sang informally on their own. Acoustic analyses revealed that infant-directed performances had fewer expressive variations in timing such as rubato and ritardando than did non-infant-directed performances. By contrast, infant-directed performances exhibited a greater dynamic range than non-infant-directed performances. Moreover, relations between pitch height and amplitude were more tightly coupled in infant-directed than in non-infant-directed performances. We interpret mothers’ liberal modulation of dynamics and their limited modulation of timing as fine-tuning to the perceptual, emotional, and informational needs of infants.

Keywords—mothers, infants, singing, timing, dynamics, fine-tuning

Mothers and other caregivers across cultures sing to infants in the course of providing care (Trehub & Trainor, 1998). Like infant-directed (ID) speech, ID singing is higher in pitch level, more emotionally expressive, and slower in tempo than non-ID singing (Trainor, 1996; Trainor, Clark, Huntley, & Adams, 1997; Trehub et al., 1997; Trehub, Unyk, & Trainor, 1993). These differences enable adult listeners to distinguish ID from non-ID performances of the same song by the same singer (Trehub et al., 1993, 1997). They may also underlie infants’ preference for ID over non-ID singing (Trainor, 1996), even in the newborn period (Masataka, 1999). Presumably, the ID singing style is part of an intuitive, multi-modal parenting repertoire that is finely tuned to infants’ abilities and needs (Papoušek, 2007).

Despite the aforementioned similarities in ID speech and singing adjustments, there are differences that arise from drastically different constraints across domains. Whereas speech, especially speech to infants, is relatively unconstrained in its pitch and temporal patterning, specific songs have prescribed pitch patterns and rhythms. In principle, mothers could use the performance features of ID speech if they created original songs for their infants. In practice, however, most mothers sing common nursery songs. As a result, they are not at liberty to expand their pitch contours by increasing the pitch range of their songs, as they do in their speech to infants (Fernald & Simon, 1984). These expanded pitch contours are thought to be largely responsible for infants’ enhanced attention and affect to ID speech relative to non-ID speech (Fernald, 1991; Fernald & Kuhl, 1987).

As is the case for the ID speech register, the primary goal of ID singing is social and emotional regulation (Shenfield, Trehub, & Nakata, 2003; Trehub & Trainor, 1998). However, mothers’ use of prosodic stress for important words and their placement of such words in utterance-final position (Albin & Echols, 1996; Fernald & Mazzie, 1991; Fisher & Tokura, 1996) imply that these adjustments have intuitive didactic as well as social-regulatory goals. Comparable didactic goals in ID singing could involve emphasis on important aspects of music such as its pitch and temporal structure.
Although the pitch and rhythmic patterns of music are highly constrained, dynamics (i.e., loudness, based on changes in sound amplitude) and some aspects of timing (e.g., tempo) have fewer constraints. As a result, variations in dynamics and timing can be used for expressive purposes. For example, higher pitches are often sung more loudly than lower pitches, although the correlation between pitch height and amplitude (in dB) is relatively modest (Scharine, 2002). If mothers intensified or exaggerated these dynamic accents in their infant-directed singing, they would simultaneously highlight the overall pitch structure of their song, its melodic (pitch) contours in particular. Accents in music, which are analogous to stress in speech, can involve higher pitch (i.e., tonic accents), greater amplitude (i.e., dynamic accents), and increased duration (i.e., agogic accents) (Handel, 1989; Patel, 2008).

Musicians often add color to their performances by expressive timing devices such as *rubato*, or slight shifts in tempo (slowing down, speeding up), and *ritardando*, or slowing down, at phrase endings (Repp, 1990). In general, listeners have little difficulty perceiving the underlying temporal structure of familiar music (e.g., its rhythm and meter) in the context of expressive timing variations. For infants, however, regular timing enhances encoding, retention, and preference for melodies and other sound sequences (Nakata & Mitani, 2005; Trehub & Hannon, 2009). If mothers were sensitive to infants’ need for temporal regularity, they should favor regular timing in their ID singing.

The purpose of the present study was to explore the expressive use of timing and dynamics in mothers’ ID singing and in non-mothers’ non-ID singing. Modulations of timing and dynamics are often considered the most important expressive devices in musical performance (Repp, 1999b), with musical structure posing fewer constraints on dynamics than on timing (Repp, 1999a).

Previous comparisons of ID and non-ID singing have featured mothers’ performances of songs in the presence and absence of their infants (Trainor, 1996; Trainor et al., 1997; Trehub et al., 1993; Trehub et al., 1997). Specifically, each mother sang the same nursery song of her choice in both contexts, which meant that different mothers sang different songs. Because mothers use a very small repertoire of songs in the course of caregiving (Trehub et al., 1997), which they perform in nearly identical fashion on different occasions (Bergeson & Trehub, 2002), their non-ID performances may be influenced by their stereotyped ID performances. Just as mothers tend to sing children’s songs when providing a sample of informal singing in the absence of their infant (Trehub et al., 1993), they may use some ID performance devices in non-ID contexts. Expressive timing differences across ID and non-ID contexts could also be obscured by variations in song choices across mothers. In short, the apparent absence of expressive timing differences in ID and non-ID singing (Trainor et al., 1997) could be attributable to the characteristic focus on mothers and on different song selections or song types (e.g., lullabies, play songs) across mothers.

In the present study, we compared performances of the same song by mothers and non-mothers who had minimal musical training. We used 10 ID performances of *Twinkle, Twinkle, Little Star* from mother-infant interactions recorded for a previous study (Nakata & Trehub, 2004). We also obtained informal non-ID renditions of the same song from an equal number of non-mothers who were comparable in age and musical training.

Notions of intuitive parenting emphasize caregiving behaviors that complement the perceptual, cognitive, and motor constraints of infants (Papoušek, 2007). Accordingly, we expected ID performances to exhibit more temporal regularity than non-ID performances because of infants’ enhanced encoding and retention of temporally regular sound sequences (Nakata & Mitani, 2005; Trehub & Hannon, 2009). For similar reasons, we expected ID performances to exhibit less *ritardando* at phrase endings than non-ID performances. Instead, we expected a greater modulation of dynamic range (i.e., extent of amplitude variations) in ID than in non-ID performances, with ID performances emphasizing the pitch structure of the music (e.g., dynamic accents on higher pitches).
METHOD

Apparatus and stimuli

All performances were recorded digitally (Sony TCD-D7 DAT recorder, Sony PC-62 lapel microphone) in a double-walled sound-attenuating room (Industrial Acoustics). ID recordings were obtained from a previous study (Nakata & Trehub, 2004) in which mothers sang songs of their choice to their 6-month-old infants in the course of informal interactions under controlled conditions. Mother and infant were seated face-to-face, very close to one another, and no observer was present in the room. Mothers were instructed to sing in their usual manner of singing to infants. Only those who sang Twinkle, Twinkle Little Star \((n = 10)\) were included in the present study. A comparison sample of non-mothers \((n = 10)\), consisting of women who were similar in age, musical background, and native language, provided a single sample of Twinkle, Twinkle, Little Star. They were asked to try, to the best of their ability, to reproduce the casual singing style they would use if they were singing on their own at home. These latter singers, who remained alone in the test room, started and stopped the recording. Obviously, conditions were different for the ID and non-ID performances. Mothers interacted with their infants prior to the recording of Twinkle, Twinkle, Little Star, sometimes by talking, sometimes by singing other songs. The non-mothers had no such interaction. Nevertheless, they were told to take as much time as they needed before beginning their recording. Thus it is unclear how many of them rehearsed their songs before the onset of recording.

Acoustic measures

Figure 1 presents a notated version of Twinkle, Twinkle, Little Star with numbered phrases and bars. The tune has a simple structure consisting of three main phrases (lines 1, 2, and 3), each of which can be subdivided into two sub-phrases. Note that the first and third lines are identical and that each of the three main phrases has 14 notes. Quarter notes dominate—12 within each line—with a longer note (equal in duration to two quarter notes) at the end of the three main phrases (corresponding to the words “are”, “sky”, and “are”) and at the boundary between sub-phrases (corresponding to the words “star”, “high”, and “star”). All of the tones of this

Figure 1. Conventional notation for Twinkle, Twinkle, Little Star, with numbered phrases and bars.
song belong to the same key, and only the first six tones of the scale (do, re, mi, fa, sol, la) are used (i.e., ti is not used). Sound spectrograms were used to identify the onset of the first vowel for each note of each singing sample. Timing, pitch, and amplitude measures were obtained for each note by means of Praat 4.3.18 speech analysis and synthesis software (Boersma & Weenink, 2005). Further analyses of these measures were conducted by means of MATLAB (Mathworks, 2007) and Systat 11 (Cranes Software, 2007).

The unit of timing was the interonset interval (IOI), as measured in ms, for each quarter note, except for the last note. Average fundamental frequency ($f_0$ in Hz) was estimated for a stable portion of the vowel corresponding to each note. Maximum amplitude (in relative dB SPL) of each vowel was the primary unit of analysis for dynamics. Amplitude was extracted with a time step of 5.3 ms and setting minimum pitch to 150 Hz. The length of the window was automatically set by the Praat software.

**RESULTS**

Figure 2 depicts average IOIs, based on quarter-note units, for the component notes of ID and non-ID versions of *Twinkle, Twinkle, Little Star*. As in previous research, ID singing was slower than non-ID singing, as reflected in longer IOIs (quarter-note units) for ID versions ($M = .48$ s, $SD = .08$) than for non-ID versions ($M = .39$ s, $SD = .08$), $t(18) = 2.66$, $p < .05$. Inspection of Figure 2 reveals smaller temporal deviations from the mean IOI (indicated by the solid line) for ID than for non-ID singing. To compare the temporal stability of both types of singing, the coefficient of variation, a normalized measure of dispersion, was used as an index of variability, with IOIs for each quarter note as the dependent measure. This coefficient was significantly smaller for ID singing ($M = .09$, $SD = .02$) than for non-ID singing ($M = .13$, $SD = .07$), $t(18) = -1.90$, $p < .05$ (one-tailed), confirming the greater temporal stability of maternal singing. A one-tailed test is justifiable on the basis of our predictions.

![Figure 2](image_url)

*Figure 2*. Average interonset intervals (IOIs) of ID and non-ID renditions, based on quarter-note units, for all notes of *Twinkle, Twinkle, Little Star*, with the exception of the final note. IOIs for half notes were rescaled as quarter notes (i.e., 50% of actual value). Phrase-endings are indicated by filled circles.
As can be seen in Figure 2, temporal variations occurred at comparable locations for ID and non-ID performances, the means for ID and non-ID singers being highly correlated, Pearson’s $r = .68$, $p < .001$. As noted, however, temporal variations were smaller for ID than for non-ID performances. To compare the extent of phrase-final lengthening in ID and non-ID singing, the mean IOI of non-phrase-ending notes (indicated by open circles in Figure 2) was subtracted from the phrase-ending notes in the first and second phrase (indicated by filled circles in Figure 2). The resulting phrase-ending deviations were significantly smaller for ID singing ($M = 2.89$ ms, $SD = 2.86$) than for non-ID singing ($M = 8.29$ ms, $SD = 5.18$), $t(18) = -2.89$, $p < .01$, reflecting mothers’ more modest use of this expressive timing device.

Degree of dynamic expressiveness was examined by means of amplitude range and by the relation between pitch height and amplitude for both types of singing. Amplitude range was greater for ID singing ($M = 15.82$ dB, $SD = 3.57$) than for non-ID singing ($M = 10.48$ dB, $SD = 4.29$), $t(18) = 4.29$, $p < .001$. Despite the greater amplitude range of ID singing, overall amplitude level was lower for ID than for non-ID singing. Although the accuracy of absolute amplitude levels is questionable in situations such as these, it was clear that non-mothers were projecting their voices in ways that mothers were not. Figure 3 depicts mean $z$-scores, or normalized amplitudes, for the six pitches (do, re, me, fa, sol, la) of the first phrase (i.e., a representative sample) from ID and non-ID performances. Relations between these $z$-scores and pitch height in the first phrase of performances were examined by means of a mixed-design analysis of variance (ANOVA) with pitch height (semitone steps corresponding to do, re, mi, fa, sol, la) as a within-subject factor and singing condition (ID, non-ID) as a between-subjects factor. There was a significant interaction between pitch height and singing condition, $F(5,90) = 2.72$, $p < .025$. Post-hoc

![Figure 3](image-url). Average $z$-scores of amplitude for the six component pitches of the first phrase of ID and non-ID renditions. Each dot represents one of the pitches.
analyses revealed a significant linear effect of pitch height in the case of ID singing \( (p < .001) \), reflecting a positive relation between pitch height and dynamics but no comparable effect for non-ID singing \( (p > .05) \).

**DISCUSSION**

In the present study, we explored variations of timing and dynamics in mothers’ infant-directed (ID) performances of *Twinkle, Twinkle, Little Star* and non-mothers’ informal, non-ID performances of the same song. In principle, playful interactions with infants in the course of singing could generate less self-monitoring of performances and, consequently, less regular timing. We found, however, that ID singing was more temporally regular than non-ID singing.

Unlike the non-ID singing, ID singing revealed relatively limited use of *ritardando*, or slowing down, at the ends of phrases, a common expressive device in music (Repp, 1990). At first glance, it might seem like mothers’ ID performances were less vocally expressive than non-mothers’ non-ID performances, perhaps because of mothers’ typical use of multimodal expressive gestures such as smiling, head movement, and touch (Trehub, 2009). This interpretation is inconsistent with the wide range of expressive vocal adjustments that mothers make in the course of speaking and singing to infants (Fernald, 1991; Trainor et al., 1997; Trehub et al., 1997). We suggest, instead, that the temporal regularity of ID singing is intuitively tuned to infants’ perceptual needs such as their enhanced processing of music with regular timing (Nakata & Mitani, 2005; Trehub & Hannon, 2009).

Regular timing may also fulfill an intuitive didactic function by increasing the transparency of various temporal features of music including the underlying pulse or beat as well as the rhythm. Strict timing regularity is obviously unnecessary for adults who listen to familiar music. In fact, performances with *tempo giusto*, or strict tempo, are often perceived as mechanical or monotonous (Gabrielsson, 2003). ID performances did not exhibit perfect regularity, or mechanical timing. As can be seen in Figure 2, the tempo variations were subtle, and they occurred at appropriate temporal positions, or those used in non-ID singing. Clearly, the ID performances were not monotonous or inexpressive. Instead, they were geared for naïve but receptive listeners who were embarking on an extended journey of musical enculturation.

In contrast to timing modulations, which were greater in non-ID than in ID singing, dynamic modulations were greater in ID than in non-ID singing. Enhanced dynamic expressiveness in ID singing was revealed in two ways. Even though ID singing was quieter overall than non-ID singing, ID singing incorporated a greater dynamic range than non-ID singing, which parallels the greater dynamic range of ID relative to non-ID speech (Fernald, 1991). Second, amplitude variations from the very first phrase of ID performances were positively related to pitch height, which is in line with emphatic stress in speech (Bolinger, 1989; Fernald & Mazzie, 1991).

Dynamic variations are used differently in speech and singing. In ID speech, dynamic accents highlight new information or key words, which often occur in privileged position at utterance endings (Albin & Echols, 1996; Fernald & Mazzie, 1991; Fisher & Tokura, 1996). Moreover, ID utterances often end on a high pitch (Fernald, 1991), unlike non-ID utterances, which typically end on a low pitch. Thus, final word or syllable stress in speech to infants frequently involves simultaneous changes in pitch (higher), duration (longer) and amplitude (louder). By contrast, songs for children (e.g., nursery songs) often end on a low pitch (e.g., the tonic) and with decreased amplitude. In fact, the tight coupling of pitch height and amplitude in the ID performances resulted in *decrescendo* or *diminuendo* (i.e., lower amplitude) as mothers ended their sung phrases.

In ID singing, greater dynamic range and singing higher pitches more loudly than lower pitches increase the salience of the component pitches and highlight the pitch contours or shape of the melody. These pitch contours may have expressive functions like those of ID speech (Fernald, 1991; Fernald & Kuhl, 1987; Trainor, Austin, & Desjardins, 2000), for example, rising contours to solicit attention and falling contours to soothe or sustain attention.
Takayuki Nakata & Sandra E. Trehub

Just as ID speech has been deemed more emotionally expressive than non-ID speech (e.g., Fernald, 1991; Trainor et al., 2000), ID singing has been deemed more emotionally expressive than non-ID singing (Trainor, 1996; Trainor et al., 1997; Trehub et al., 1997). These differences may be overstated in speech as well as singing. Typical speech comparisons involve affectively positive ID speech and affectively neutral non-ID speech (e.g., Fernald & Simon, 1984), which have drastically different intentions—emotional regulation in one case and information transmission in the other. When the expressive intentions of ID and non-ID speech are comparable, as in expressions of love or fear, the resulting prosodic differences are minimal (Trainor et al., 2000). In fact, infants seem to prefer affectively positive to affectively neutral speech, regardless of its ID or non-ID status (Singh, Morgan, & Best, 2002).

Whereas ID speech and singing to pre-verbal listeners are relatively similar in their communicative intentions (i.e., emotional regulation), non-ID speech and singing are not. Samples of non-ID speech in the available literature are typically drawn from emotionally neutral communicative exchanges between adults (e.g., Fernald & Simon, 1984) or from scripted portrayals of such exchanges (e.g., Trainor et al., 2000; Werker & McLeod, 1989). By contrast, samples of non-ID singing usually feature mothers singing with no visible audience in a laboratory setting (Trainor, 1996; Trehub et al., 1997; Trehub et al., 1993) or, as in the present study, non-mothers doing so. Presumably, musically untrained individuals who choose to sing on their own are motivated by emotional self-regulation or pleasure. Thus, there is no reason to expect their singing to be less expressive than ID singing, simply different because of contrasting goals.

Despite the unnatural context of non-ID singing in the present study, which undoubtedly inhibited the untrained performers, such singing revealed some expressive timing variations that are evident in the performances of musicians (Repp, 1990). The results of the present study are at odds with the generalization that ID singing is simply more expressive than non-ID singing. What emerged, instead, was that ID and non-ID singing differed in their expressive style, with variations in dynamics making a greater contribution to the expressiveness of ID singing and variations in timing making a greater contribution to the expressiveness of non-ID singing. We contend that the modest timing modulations and robust dynamic modulations in ID singing reflect fine-tuning to infants’ perceptual, emotional, and informational needs. The timing modulations in non-ID singing may reflect adherence to cultural conventions rather than emotional self-regulation. Obviously, the laboratory context is a challenging context for exploring expressive variations in informal, solo singing. We suggest, however, that singing for one’s own pleasure is a topic worthy of further research, as is ID singing.

REFERENCES


**AUTHOR NOTES**

Funded by the Social Sciences and Humanities Research Council of Canada.
BIOGRAPHIES

Takayuki Nakata received his doctoral degree in psychology from Texas Christian University and completed a postdoctoral fellowship at the University of Toronto. Currently, he is an Associate Professor of Psychology at Future University Hakodate in Japan. His research interests include mother's singing and speech to infants and the perception and production of music and speech prosody by deaf children with cochlear implants.

Sandra Trehub, who obtained her doctorate from McGill University in 1973, is Professor Emeritus of Psychology at the University of Toronto and Visiting Professor at the International Laboratory for Brain, Music and Sound Research (BRAMS) at the Université de Montréal. She has published extensively on infants’ perception of music and parents’ songs to infants. Although most of her research is conducted in her laboratory, she sometimes travels to remote regions of the world to observe mothers’ musical interactions with infants.