Children's Expression of Emotion in Song
Mayumi Adachi and Sandra E. Trehub
_Psychology of Music_ 1998; 26; 133
DOI: 10.1177/0305735698262003

The online version of this article can be found at:
http://pom.sagepub.com/cgi/content/abstract/26/2/133

Published by:
SAGE
http://www.sagepublications.com

On behalf of:
Society for Education, Music and Psychology Research

Additional services and information for _Psychology of Music_ can be found at:

Email Alerts: http://pom.sagepub.com/cgi/alerts
Subscriptions: http://pom.sagepub.com/subscriptions
Reprints: http://www.sagepub.com/journalsReprints.nav
Permissions: http://www.sagepub.co.uk/journalsPermissions.nav
Citations http://pom.sagepub.com/cgi/content/refs/26/2/133
Children’s Expression of Emotion in Song

MAYUMI ADACHI
Faculty of Education and Human Sciences, Yamanashi University, 4-4-37, Takeda, Kofu, Yamanashi 400-8510, Japan.
E-mail: adachi@grape.kkb.yamanashi.ac.jp

and

SANDRA E. TREHUB
Department of Psychology, University of Toronto at Mississauga, Mississauga, Ontario, Canada, L5L 1C6.
E-mail: trehub@credit.erin.utoronto.ca

Abstract
Children 4–12 years of age (N = 160) were recorded (audio and video) as they sang two versions of a familiar song, once in an attempt to make an adult listener happy and once to make her sad. Coding of gestural, vocal, linguistic and musical devices revealed that children used all of these means to portray contrastive emotions. Regardless of age or singing skill, children relied primarily on expressive devices used in interpersonal communication (e.g. tempo, facial expression) and made relatively little use of music-specific devices (e.g. legato). Moreover, they used a greater variety of expressive devices in their sad performances than in their happy performances. Finally, age-related changes reflected the influence of maturity, socialisation and musical knowledge.

The rudiments of emotional communication are apparent in the second year of life (Bretherton, Fritz, Zahn-Waxler and Ridgeway, 1986; Bretherton, McNew and Beeghly-Smith, 1981; Yarrow and Waxler, 1977). By 20 months of age, children reliably decode conventional facial expressions of happiness and sadness (Bretherton et al., 1981). When confronted with expressions of dejection, they offer comfort (Borke, 1971; Bretherton and Beeghly, 1982; Bretherton et al., 1986; Bretherton et al., 1981; Hoffman, 1981; Zahn-Waxler and Radke-Yarrow, 1982, cited in Bretherton et al., 1986). By two years of age, children label how they or others feel (Bretherton and Beeghly, 1982; Bretherton et al., 1986), exhibiting some understanding of the antecedent events that elicit particular feelings (Bretherton and Beeghly, 1982; Radke-Yarrow and Zahn-Waxler, 1973, cited in Bretherton et al., 1986). Once they begin to tell stories, pre-schoolers use colourful vocal tones to keep their audience emotionally engaged (Reilly, 1992). In other words, pre-school children use conventional means of emotional expression in their everyday communication. The prevailing wisdom is that children acquire knowledge of emotional expression through interaction with their care-givers, such knowledge laying the foundation for general social competence (Bretherton, et al., 1981; Cole, 1985; Lewis and Michalson, 1983; Ratner and Stettner, 1991; Saarni, 1985; 1993). Despite the wealth of information on the development of emotional communication in everyday life (e.g. Bretherton et al., 1986; Cole,
1985; Harris, 1989; 1993; Saarni, 1993), relatively little is known about emotional communication in music (Juslin, 1997a) and even less about its development.

Emotional communication by means of music begins in the early months of life. For example, mothers sing to their infants to soothe or amuse them (Trehub and Schellenberg, 1995; Trehub and Trainor, 1998; Trehub, Unyk and Trainor, 1993). Some of the features that distinguish mothers’ songs to infants from other versions of the same song are higher pitch, slower tempo and a more emotive voice quality (Trainor, 1996; Trehub and Trainor, 1998; Trehub et al., 1997). Even pre-school children make similar vocal alterations when singing to their infant siblings (Trehub, Unyk and Henderson, 1994). Whether pre-school children acquire this emotive style of singing from their mothers or whether it is an unlearned affectional response to infant visual (Sternglanz, Gray and Murakami, 1977) and vocal (Bloom and Lo, 1990) cues remains to be determined. What is also unclear is whether children can communicate emotion in song, as they do in speech.

Music functions, among other things, as a medium of communication (Jones and Holleran, 1992; Sessions, 1967), the information communicated being primarily emotional (Aiello, 1994; Meyer, 1956). Adults are generally successful in decoding the emotional intentions of composers, as in musical selections such as the Funeral March in Beethoven’s Symphony No. 3, which expresses sadness, despair and grief and Nicolai’s Merry Wives of Windsor, which expresses joy and triumph (Hampton, 1945). Hill, Kamenetsky and Trehub (1996) determined that Baroque composers used the Ionian (major) mode for hymns with positive content (salvation) and the Phrygian mode for those with negative content (condemnation) and that contemporary adults and children make comparable associations between text (positive, negative) and mode (Ionian, Phrygian).

Specific musical elements function as emotional cues. For example, the major and minor modes are linked to positive and negative affect, respectively (Crowder, 1984; 1985; Hevner, 1935; 1936; Wedin, 1972). Moreover, high pitch, fast tempo, loud sound and staccato articulation are associated with happiness, in contrast to low pitch, slow tempo, soft sound and legato articulation, which are associated with sadness (Heinlein, 1928; Hevner, 1937; Juslin, 1997b; Rigg, 1940; Wedin, 1972).

In addition to expressive means within a piece of music, expressive devices in the performance can intensify or alter the emotional interpretation (Hampton, 1945; Sessions, 1967). Performers manipulate the expressiveness of their performances (Davidson, 1993; Davidson and Dawson, 1995; Kendall and Carterette, 1990), making it possible to project different emotional qualities onto a piece of music in a way that is readily decoded by musically trained or untrained adults (Behrens and Green, 1993; Gabrielsson and Juslin, 1996; Juslin, 1997a; Ohgushi and Hattori, 1996; Senju and Ohgushi, 1987). For example, instrumental performers convey happiness by playing the notes of a musical piece faster, louder, more brightly and distinctly, with a fast and light vibrato; they convey sadness by playing the notes slower, softer and more smoothly, with a slow and deep vibrato (Gabrielsson and Juslin, 1996; Juslin, 1997a).

Successful emotional communication between composers and performers, on the one hand, and naïve adult listeners, on the other, likely implicates conventional
as well as natural (i.e. untrained) means of expressing and interpreting particular emotions. More adult-like interpretations of music in older than in younger children (Cunningham and Sterling, 1988; Dolgin and Adelson, 1990) are suggestive of culture-specific means of emotional expression that are not yet mastered by young children. Although there is some information about children’s perception of emotion in music, there is no information about their ability to convey particular emotions in their own musical performances.

We examined children’s attempts to express particular emotions by means of their informal musical performances. We chose singing as a performance medium because of the early onset of spontaneous singing and its importance in early childhood (Davidson, McKernon and Gardner, 1981; Dowling, 1984; Kelley and Sutton-Smith, 1987; McKernon, 1979; Moorehead and Pond, 1941; 1942; Umezawa, 1990). Happiness and sadness were chosen as the target emotions because these basic emotions are discernible and reproducible by children as young as 2 years of age (Borke, 1971; Bretherton et al., 1981). A highly familiar song (Twinkle, Twinkle, Little Star, The ABC Song or Baa, Baa, Black Sheep) was selected to lessen the cognitive demands on children and facilitate their expression of emotion. Children were asked to sing the song in a way that would make the listener feel happy (or sad). Subsequently, they were asked to sing it so that the listener would feel sad (or happy). Audio and video recordings permitted access to vocal and gestural means of emotional expression.

Emotional cues in vocal performances go well beyond the usual auditory cues, involving highly controlled, stylized behaviours that include bodily stance, facial expression and a range of vocal effects (Pantaleoni, 1985). In some instances, audiences have been found to interpret vocalists’ emotional intentions more accurately from soundless video materials than from audio materials alone (Oghushi and Hattori, 1996).

Most research on emotional expression in children has focused on facial cues. Children as young as 3 months of age smile and frown during their interactions with adults (Emde, Kligman, Reich and Wade, 1978). By the pre-school period, they respond to positive thoughts with a smile and to negative thoughts with a downcast expression (Masters, Barden and Ford, 1979). They can simulate facial expressions of happiness and sadness even when they feel otherwise (Borke, 1973). Moreover, they correctly decode the facial expressions of story characters (Burns and Cavey, 1957; Gross and Harris, 1988) and portray a variety of facial emotions during pretend play (Bretherton et al., 1981; Wolf, Rygh and Altschuler, 1984). Adult-like decoding of happiness and sadness from postural and movement cues (DeMeijer, 1989; 1991) is not achieved until 8 years of age (Van Meel, Verburgh and DeMeijer, 1993). In some circumstances, however, children as young as 4 and 5 years of age interpret posed or pictorial depictions of happy and sad expressions like adults do (Boone and Cunningham, 1997; Hobson, 1986).

Children’s use of expressive cues in their musical performances is likely to differ as a function of age, gender and musical ability. For example, pre-schoolers rely primarily on vocal affect and school-age children on content to convey emotion in their stories (Reilly, 1992). Although pre-schoolers commonly react to music by spontaneous body movement (Metz, 1989), older children are inclined to suppress such movement (Cohen, 1980/1981). There is much evidence to suggest that girls are better encoders of emotion than are boys (Buck, 1975; Carlson, Gantz and
Masters, 1983; Wolf et al., 1984), perhaps because boys are socialised to suppress emotional displays and girls are encouraged to react expressively and empathetically to others (Brody, 1985; Brody and Hall, 1993; Fuchs and Thelen, 1988; Lewis and Michalsen, 1983; Mussen, 1969). Finally, singing skill is likely to have an impact not only on the overall expressiveness of a performance but on the expressive devices used to achieve specific emotional goals. Presumably, singers with greater skill would use more conventional musical devices than those with lesser skill.

Method

Participants

The participants were 224 children 4–12 years of age: 47 4-year-olds, 36 5-year-olds, 31 6-year-olds, 27 7-year-olds, 25 8-year-olds, 18 9-year-olds, 10 10-year-olds, 19 11-year-olds and 22 12-year-olds. Children were recruited from a large sample of families who had volunteered to participate in university research on child development. Children were excluded from the final sample because of equipment/technical problems (n = 7), colds on the test day (n = 2), inability to do the tasks (n = 12), or failure to complete all of the singing tasks (n = 43). The final sample of 160 children consisted of 40 children at each of four age-levels: 4–5 (M = 4.45), 6–7 (M = 6.55), 8–10 (M = 8.75) and 11–12 (M = 11.55) years of age. Half of the children at each age level were “good” singers in the sense that they accurately reproduced all intervals and maintained the key while singing a familiar song with and without lyrics; the remaining children were “ordinary” singers. The proportion of girls and boys was roughly equivalent for ordinary singers at all age levels and for good singers 6 to 7 years of age. Because of the difficulty of finding boys who were skilled singers at younger and older age-levels, girls outnumbered boys among good singers by a 4:1 (4–5 years) or 3:1 (8–10 and 11–12 years) ratio.

Apparatus

All testing was conducted on university premises in a double-wall sound-attenuating booth (IAC#110766) equipped with audio and video recording systems (SONY TCD-D7 digital audio tape recorder, SONY ECM-959A microphone, Panasonic WV-1500 video camera).

Procedure

Prior to entering the test booth, children were told that they would be asked to sing some songs and tell a story. They were also assured that we were merely playing a game together; there was no right or wrong way of performing any task. Each child was seated facing the experimenter and the entire session was recorded (separate audio and video). The session began by asking the child to sing Twinkle, Twinkle, Little Star (“Twinkle”), The ABC Song (“ABC”), or Baa, Baa, Black Sheep (“Baa”) – whichever was more familiar – with and without lyrics. As noted, any child who reproduced intervals accurately and remained within one key in both versions was classified as a “good” singer; otherwise, the child was classified as an “ordinary” singer. To orient the child to the expressive singing tasks that followed, children 6–12 years of age were asked to tell a story to make the experimenter happy (for half of the children) or sad (for the other half). Children 4–5 years of age made up a story based on a drawing of a happy
bunny (for half of the children) or a sad bunny (for the other half). Subsequently, children of all ages were asked to make up two songs (free singing task) – one with words and the other without words – to make the experimenter feel as happy as (or as sad as) she had felt for the story. Finally, in a restricted singing task – the focus of the present report – children were asked to sing the familiar song sung earlier (i.e. “Twinkle,” “ABC” or “Baa”) in two emotional contexts (with order counterbalanced): (a) to make the experimenter feel happy (or sad), as in the previous tasks, and (b) to induce the opposite emotion. The experimenter maintained a neutral demeanour with children aged 6 or older. For the 4- and 5-year-olds, the experimenter manipulated a bear puppet to portray the target emotion. Children received no explicit guidance for their performances of the familiar song. In principle, they were free to manipulate the musical attributes of the song, the lyrics, as well as vocal and gestural aspects of their performance. Children were uniformly co-operative and attentive for all of the tasks. On average, they completed the entire procedure in 10–15 minutes, their contrasting performances of the familiar song requiring 3–4 minutes.

Data Scoring

Children’s gestural devices were extracted from the video recordings and their vocal, linguistic and musical devices from the digital audio recordings. Although most devices were coded categorically (e.g. present or absent), three (i.e. tempo, dynamics and pitch) were measured instrumentally.

Instrumental analyses. Tempo differences across emotional contexts were calculated, first, by measuring the number of beats per minute (treating each two-syllable beat as a unit) near the beginning of the song and, subsequently, generating a deviation score (tempo of happy version minus that of sad version). Differences in dynamics (loudness) across contexts were calculated, first, by measuring the peak amplitude of the first available vowel in each song and, subsequently, converting the resulting voltages to deviation scores in dB SPL by means of the formula, \(20 \log_{10} \left( \frac{v(\text{happy})}{v(\text{sad})} \right)\) (where \(v\) denotes voltage). Pitch differences were obtained by measuring the fundamental frequency \((F_0)\) of the first available vowel in each song and converting the \(F_0\) from happy and sad versions to deviation scores (in semitones) by means of the formula, \(12 \log_2 \left( \frac{F_0(\text{happy})}{F_0(\text{sad})} \right)\).

Categorical coding. Facial expression (frown, neutral, smile), eye position (straight ahead, up, down, side), head position (upright, downward, tilted, turning, other), head movement (present, absent), bodily posture (upright, forward, backward) and body movement (present, absent) were coded at the onset of the sung performance. Definitions of facial expressions were drawn from Izard (1971) and Blairton Jones (1971). Leg movement (present, absent) was coded for the first 16 beats (i.e. “Twinkle, … what you are”, “ABC … P” and “Baa, … full”) in each emotional context. If the nature of leg movement during the sung performance was similar when the child was not singing, it was considered habitual and not coded as leg movement.

Vocal devices were coded on a binary basis (present, absent). In two instances, breathy voice and nasal voice, an independent rater judged whether the singing sample of a child (first 16 beats of each sung performance) was breathy or nasal, as in the examples provided. The category, unique voice, was used to characterise unusual vocal manipulations throughout a song that generated a piercing or falsetto
voice quality, or that involved squeaky, throaty, laughing, crying, mumbling, or animal sounds.

Four linguistic devices were coded on a binary basis (present, absent) by comparing the happy and sad versions with the version used for initial classification. Emotion-evoking words were coded as present when words such as “happy”, “good”, “love”, “good-bye”, “cry” and “sad” replaced or were added to the original lyrics. Paralinguistic strategy referred to the use of negation or affective prosody such as pretend giggle or sniff. Word removal was coded when at least one word was removed from the lyrics used in the classification task. The presence of new lyrics was coded when most of the words of the song were changed.

Musical devices consisted of manipulations of rhythm, articulation and melody. The scoring of rhythmic and melodic manipulations involved comparisons of the happy and sad versions with the classification version. Scoring was binary (present, absent) for all but melodic change, which involved four sub-categories. Dotted or uneven rhythm was coded as present when tones of equal duration were replaced by a combination of longer and shorter tones that left the original metrical accent intact (see Figure 1). Augmentation occurred when tone duration was extended, whereas diminution occurred when tone duration was shortened (Figure 1). Syncopation was present when the position of a metrical accent was moved and the duration of tones changed (Figure 1).

Articulatory manipulations included legato (at least three consecutive tones glued together so that consonant onset was subdued), non-legato (at least three

<table>
<thead>
<tr>
<th>Original Rhythm Examples</th>
<th>Altered Rhythm Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dotted/Uneven</td>
<td></td>
</tr>
<tr>
<td>Augmentation</td>
<td></td>
</tr>
<tr>
<td>Diminution</td>
<td></td>
</tr>
<tr>
<td>Syncopation</td>
<td></td>
</tr>
</tbody>
</table>

Fig. 1

Examples for four types of rhythmic manipulations: (1) dotted or uneven rhythm, (2) augmentation, (3) diminution, and (4) syncopation.
consecutive tones clearly separated), accent (any tone(s) in the tune or phoneme(s) in the lyrics clearly emphasized) and tenuto (any tone(s) or phoneme(s) sustained for purposes of emphasis). Melodic manipulations included speech-like melody (melody produced with a speaking rather than singing voice as in monologue, chant, or cheer), suppressed-contour (magnitude of contour change reduced while singing) and portamento (two consecutive pitches connected by modulation or gliding). In addition, the degree of melodic change was assessed as extreme (resulting in a new tune), substantial (at least two-thirds of the melody modified), partial (isolated intervallic change), or none (melody unchanged).

Interrater agreement was calculated by having a second rater code 20% of the corpus at each age level for each device. High interrater agreement, ranging from .86 to 1.0, was achieved for most devices. Lower interrater agreement occurred in the following instances: leg movement of 8–10-year-olds (.81), diminution of 8–10-year-olds and 11–12-year-olds (.81), legato of 11–12-year-olds (.81), tenuto of 6–7-year-olds (.81) and bodily posture of 11–12-year-olds (.79). The poorest agreement was found for dotted rhythm of 6–7-year-olds (.75) and for dotted rhythm, augmentation, diminution and tenuto (.75) on the part of 4–5-year-olds. Disagreements between raters were resolved by discussion and, in some cases, by consultation with a third, independent rater. In the case of legato, non-legato and tenuto, this discussion and consultation resulted in minor revisions of the coding scheme and recoding of the entire data set.

Results

Due to unequal numbers of boys (n = 23) and girls (n = 57) in the category of good singers, the factor of gender was examined by means of supplementary analyses. Categorical and non-categorical data were analysed separately. Analyses of non-categorical data are presented first, followed by analyses of categorical data.

Tempo, Dynamics and Pitch

Preliminary analyses revealed no gender differences in children’s manipulations of tempo, dynamics or pitch. Preliminary review of the data also showed that children’s creation of new lyrics or new tunes did not influence their manipulations of tempo, dynamics or pitch. Table 1 indicates the mean differences in tempo, dynamics and pitch between emotional contexts for eight sub-groups of children (good and ordinary singers at four age-levels). The uniformly positive values indicate that children, regardless of age and singing ability, sang their songs faster, louder and at a higher pitch level for their happy versions than for their sad versions. As can be seen in Table 1, the magnitude of these manipulations differed as a function of age and/or singing ability. A MANOVA (using Pillai’s trace, one of the most robust multivariate tests) indicated a significant effect of age on tempo, dynamics and/or pitch, F (9, 456) = 2.356, p = .013; no other effects were significant. Subsequent univariate tests revealed that the effect of age was significant only for tempo, F (3, 152) = 3.21, p = .025. To identify specific differences in tempo manipulations across age levels, post-hoc deviation contrasts (95% confidence intervals) were performed by comparing the mean of each age level to the overall mean (see Table 2). The results indicated that children 8–10 years of age used significantly greater tempo changes across emotional contexts than did children.
of other ages. By contrast, children 4–5 years of age manipulated tempo significantly less than other age-groups.

**Gestural, Vocal, Linguistic and Musical Devices**

Preliminary analyses revealed eight devices associated with gender. Compared to girls, boys more often frowned, \( \chi^2 (2, N = 320) = 8.88, p = .012 \), leaned backward, \( \chi^2 (2, N = 320) = 15.184, p = .0005 \), used a unique voice, \( \chi^2 (1, N = 320) = 14.058 \) (corrected for continuity), \( p = .0002 \) and speech-like melody, \( \chi^2 (1, N = 320) = 8.127, p = .004 \). By contrast, girls used breathy voice, \( \chi^2 (1, N = 320) = 15.547, \ p = .0001 \), *legato*, \( \chi^2 (1, N = 320) = 6.74, p = .009 \) and *portamento*, \( \chi^2 (1, N = \)
320) = 6.681, p = .01, more frequently than boys. These tendencies were consistent across context, age level and singing ability for all but portamento. Portamento was related interactively to gender and age. The effects of gender on other devices were largely independent of the effects of other factors. Accordingly, data from boys and girls were pooled for the remaining analyses (unless specified).

Data for each categorical device were arranged in a multiple-contingency table consisting of three independent variables: emotional context (happy, sad), age (four levels) and singing ability (ordinary, good). Because there were no clear expectations of the devices that would be associated with emotional context, age, or singing ability, hierarchical loglinear analyses were conducted to identify the most important associations between variables. Backward elimination of insignificant terms was used because it is considered superior to the forward method in hierarchical loglinear models (Benedetti and Brown, 1978). The criterion for removal of a particular term from the loglinear model was the least significant change in the likelihood ratio chi-square evaluated at α = .01.* Associations were determined between emotional context and the following devices: facial expression, eye position, head position, nasal voice, unique voice, tenuto, speech-like melody, suppressed contour and melodic change. Seven devices were associated with age-level: eye position, leg movement, emotion-evoking words, new lyrics, legato, non-legato and portamento. Portamento was the only device associated with singing ability. The use of body movement was associated with age and singing ability in an interactive manner.

Because we were interested in the differential use of expressive devices across emotional contexts, age-levels, or singing ability, we used correspondence analysis to examine how each of the devices was used in relation to each factor. Correspondence analysis transforms information from a two-way contingency table to scores plotted in one or more dimensions, yielding similarities and differences between categories either within or across variables (Agresti, 1990; Greenacre, 1981; 1993; van der Heijden, de Falguerolles and de Leeuw, 1989; van der Heijden and de Leeuw, 1985). The amount of information represented in a particular dimension (α) is indicated by the proportion of the chi-squared value that is decomposed in that dimension, calculated by the inertia for dimension α (λ^2_α) divided by the total inertia (λ^2_T). Because total inertia is equivalent to Pearson’s goodness-of-fit chi-squared statistic divided by the total number of observations (N = 320), the statistical significance of associations between variables can be determined by means of Pearson chi-square (Greenacre, 1993).

Devices related to emotional contexts. Figure 2 illustrates a one-dimensional plot of the relations between categories derived from correspondence analysis for facial expression and emotional context. The ratio of the inertia for this dimension to the total inertia (λ^2_1/λ^2_T) was 1.0, reflecting the fact that the entire relation between these variables is represented by this plot. Notice that scores for the two emotional contexts are split into positive and negative values. Categories whose scores are close to 0 are less important than those whose scores are farther from 0. Moreover, categories whose contribution to the dimension (indicated in parentheses in Figure 2) is very low (say, less than .1) are less important than

*For similar exploratory uses of hierarchical loglinear analysis for multiple contingency tables, see van der Heijden and de Leeuw (1985).
categories whose contribution is high. Based on these criteria, the relation between facial expression and emotional context was as follows: children smiled in their happy versions and frowned in their sad versions, $\chi^2 (2, N = 320) = 41.396$, $p < .00001$. Relations between other variables and emotional context were determined in a similar manner, with comparable one-dimensional plots. Children looked straight ahead and held their head upright in the happy context but exhibited a downward posture and regard in the sad context, $\chi^2 (3, N = 320) = 13.647$, $p = .003$ for eye position and $\chi^2 (4, N = 320) = 16.019$, $p = .003$ for head position. Children tended to change the melody in their sad version, but not in their happy version, $\chi^2 (2, N = 320) = 28.016$, $p < .00001$. The use of other devices was associated with the sad context, $\chi^2 (1, N = 320) = 13.278$ (corrected for continuity), $p = .001$ for nasal voice; $\chi^2 (1, N = 320) = 10.394$ (corrected for continuity), $p = .001$ for unique voice; $\chi^2 (1, N = 320) = 14.641$ (corrected for continuity), $p = .0001$ for tenuto; $\chi^2 (1, N = 320) = 11.621$ (corrected for continuity), $p = .0007$ for speech-like melody and $\chi^2 (1, N = 320) = 10.982$ (corrected for continuity), $p = .0004$ for suppressed contour.

Devices related to age-level. Figure 3 shows a two-dimensional plot representing relations between eye position and age-level. The proportion of the relation explained by dimension 1 ($\lambda^2_1/\lambda^2_T$) was .766 and the proportion explained by dimension 2 ($\lambda^2_2/\lambda^2_T$) was .138; thus, approximately 91% of the relations between eye position and age-level are depicted in this plot. Contributions of 4–5-year-olds and 8–10-year-olds as well as those of “straight ahead” and “up” were mainly to dimension 2, representing only 14% of all information. Thus, the involvement of these categories was minimal. Nevertheless, there were important relations between eye position and age-level: 6–7-year-olds tended to look sideways (i.e. left or right) and 11–12-year-olds tended to look down, $\chi^2 (9, N = 320) = 24.052$, $p = .004$. All other results for variables related to age level were plotted one-dimensionally, which explained 100% of the relations, as in the example provided in Figure 2. Note that portamento was examined against an interactive variable, age-level and gender, on the basis of preliminary inspection of the data. Leg movement was used by 4–5-year-olds, $\chi^2 (3, N = 320) = 15.81$, $p = .001$. Lyrics were changed by 6–7-year-olds, $\chi^2 (3, N = 320) = 28.022$, $p < .00001$ for emotion-evoking words and $\chi^2 (3, N = 320) = 23.684$, $p = .00003$ for new lyrics.
Two-dimensional plot representing relations between eye position and age
level, as obtained from correspondence analysis. The proportion of relations
explained by dimension 1 is .766, and that explained by dimension 2 is .138.
The first of the two values shown in parentheses indicates the contribution
of each category to dimension 1; the second value indicates the contribution
of the category to dimension 2.

*Legato* was used by 6–7-year-olds and 11–12-year-olds, $\chi^2 (3, N = 320) = 15.162,
$p = .0017$, whereas *non-legato* was used by 4–5-year-olds, $\chi^2 (3, N = 320) = 32.637,
$p < .00001$. *Portamento* was used by 11–12-year-old girls, $\chi^2 (7, N = 320) = 34.825,
p = .00001$.

*Devices related to singing ability.* Relations between *portamento* and singing
ability and between body movement and an interactive variable, age level and
singing ability, were plotted one-dimensionally, which accounted for 100% of
the information. *Portamento* was used by good singers, $\chi^2 (1, N = 320) = 11.621
(corrected for continuity), $p = .0007$ and body movement by 8–10-year-old ordinary
singers, $\chi^2 (7, N = 320) = 24.57$, $p = .001$. 

Specifying Predictor Variables for Children's Intended Emotions

The joint use of loglinear analyses and correspondence analyses allowed us to specify each device associated with emotional context, age-level and/or singing ability. Of all of the categorical devices identified, nine expressive devices were related to emotional context, that is, to children's emotional intentions. This information, however, does not reveal the relative contribution of these expressive devices to the intended emotion. Presumably, some devices would make more important and more distinctive contributions to emotional expression than would others. Moreover, two or more devices might provide redundant information. To identify subsets of devices that differentiated children's happy performances from their sad performances, we conducted step-wise variable selection using logistic regression.* Because the model takes categorical and non-categorical variables, all 12 devices associated with emotional context were included in the analysis. For non-categorical data, values from the original measurement were used. The likelihood ratio chi-square test ($\alpha = .05$) was used as a criterion for entry or removal of a term. Table 3 shows the results of variable selection. Improvement in chi-square indicates the relative importance of variables in classifying the emotional context of the sung performances.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Improvement ($\chi^2$)</th>
<th>d.f.</th>
<th>$p$</th>
<th>Correct Classification (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tempo</td>
<td>61.035</td>
<td>1</td>
<td>&lt; .001</td>
<td>70.63</td>
</tr>
<tr>
<td>Initial Facial Exp.</td>
<td>28.851</td>
<td>1</td>
<td>&lt; .001</td>
<td>71.25</td>
</tr>
<tr>
<td>Melodic Change</td>
<td>9.488</td>
<td>1</td>
<td>.002</td>
<td>75.31</td>
</tr>
<tr>
<td>Dynamics</td>
<td>10.433</td>
<td>1</td>
<td>.001</td>
<td>75.94</td>
</tr>
<tr>
<td>Nasal Voice</td>
<td>7.884</td>
<td>1</td>
<td>.005</td>
<td>77.50</td>
</tr>
<tr>
<td>Speech-like Melody</td>
<td>6.979</td>
<td>1</td>
<td>.008</td>
<td>77.81</td>
</tr>
<tr>
<td>Melodic Change</td>
<td>3.774</td>
<td>1</td>
<td>.052</td>
<td>75.94</td>
</tr>
<tr>
<td>Suppressed Contour</td>
<td>9.272</td>
<td>1</td>
<td>.002</td>
<td>78.44</td>
</tr>
<tr>
<td>Initial Eye Position</td>
<td>5.915</td>
<td>1</td>
<td>.015</td>
<td>78.13</td>
</tr>
</tbody>
</table>

Tempo, by itself, correctly classified 70.63% of sung performances into the intended emotional context. Theoretically, 50% of performances can be classified correctly by chance. The percentage above chance level (i.e. 20-63%) represents the contribution made by tempo, the most effective predictor variable. Correct classification increased cumulatively by adding facial expression (0.6%), melodic change (0.06%), dynamics (0.63%), nasal voice (1.56%) and speech-like melody (0.31%). Subsequently, melodic change was removed from the model because

---

*In logistic regression, independent variables can be categorical, non-categorical, or a combination of these. The model can predict membership of a particular case by means of information obtained from a set of independent variables. Step-wise variable selection can identify subsets of independent variables that are effective predictors.
of the negative improvement in its chi-square value; perhaps addition of speech-like melody made melodic change relatively unimportant. Correct classification was decreased by the removal of melodic change (−1.87%), but it was subsequently increased by 2.5% with suppressed contour. The last selected variable, eye position, resulted in a slight decrease in correct classification (−0.31%). Further examination of the classification table revealed that, without eye position, the model classified happy performances 4% less accurately than sad performances. By adding eye position, however, the accuracy of classification for happy and sad performances became more homogeneous. The seven-variable model, including eye position, correctly classified 77.5% of happy performances and 78.75% of sad performances, with an overall accuracy of 78.13%. Table 4 shows the correlation matrix of parameter estimates based on the seven-variable model. Because the correlations are very small, these variables are unlikely to contribute redundant information. Thus, the seven variables can be considered the most effective predictors of children’s happy and sad performances of a familiar song.

<table>
<thead>
<tr>
<th></th>
<th>Initial Eye</th>
<th>Suppressed</th>
<th>Nasal</th>
<th>Speech-like</th>
<th>Dynamics</th>
<th>Initial Facial</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tempo</td>
<td>−119</td>
<td>−088</td>
<td>017</td>
<td>−153</td>
<td>−016</td>
<td>−049</td>
</tr>
<tr>
<td>Initial Facial</td>
<td>−011</td>
<td>−098</td>
<td>−019</td>
<td>030</td>
<td>060</td>
<td></td>
</tr>
<tr>
<td>Dynamics</td>
<td>072</td>
<td>−051</td>
<td>−113</td>
<td>−221</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Speech-like</td>
<td>064</td>
<td>036</td>
<td>069</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nasal Voice</td>
<td>035</td>
<td>015</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Suppressed Cont.</td>
<td>072</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Discussion**

The present investigation revealed that children 4–12 years of age can portray happiness and sadness contrastively through their sung performances of a familiar song. We identified approximately 30 expressive devices – facial, gestural, vocal and musical – one-third of which were associated with particular emotions across age, gender and singing ability. For example, children typically smiled and looked ahead for their happy renditions; they frowned and looked downward for their sad renditions. These expressions, which are highly conventional in communicating happiness and sadness, are consistent with children’s interpretation of the emotive gestures of others (Boone and Cunningham, 1997; Hobson, 1986; Masters et al., 1979).

Boys’ and girls’ predominant vocal means of expressing emotion at all age-levels was to sing faster, louder and at a higher pitch level when portraying happiness rather than sadness. This finding parallels conventional practices in speech (Scherer, 1986) and music (Gabrielsson and Juslin, 1996; Heinlein, 1928; Hevner, 1937; Rigg, 1940; Wedin, 1972). Children’s use of *tenuto* (i.e. sustained emphasis of tones) to express sadness was functionally analogous to the “slow and deep vibrato” of trained instrumentalists (Gabrielsson and Juslin, 1996). Nasal voice, as in some
children’s sad renditions, commonly occurs in adults’ expression of unpleasant affect (Scherer, 1986). Another way in which some children portrayed sadness was by means of a unique voice – whispering, throaty, or mumbling – which shares properties of adults’ “lax” voice (i.e. whispery, muffled, imprecise articulation) in sad contexts (Scherer, 1986). Children who expressed sadness by means of speech-like melody and suppressed contour, both of which involve a reduced pitch range, were using common devices for the expression of sadness in speech (Saarni, 1982; Scherer, 1981). Although there are no adult parallels to children’s manipulation of melody in their sad versions, this option is generally excluded from adults’ expressive performances (Gabrielsson and Juslin, 1996; Juslin, 1997a). In short, children’s means of expressing happiness and sadness in song reflected conventional means of expressing these emotions in speech and music.

Some devices for expressing happiness and sadness, such as manipulations of tempo, dynamics, pitch level and pitch range, are common to music and speech. Others, such as manipulations of articulation and mode, are music-specific. In the present study, children at all age-levels capitalised primarily on domain-general devices (e.g. tempo, dynamics, pitch) and relatively infrequently on music-specific devices such as legato and non-legato (or staccato) or major and minor modes (Crowder, 1984; 1985; Gabrielsson and Juslin, 1996; Hevner, 1935; 1936; Juslin, 1997a; Wedin, 1972). Children’s predominant use of domain-general expressive elements was confirmed by the seven devices that correctly classified roughly 80% of the sung performances – tempo, facial expression, speech-like melody, suppressed contour, nasal voice, eye position and dynamics. Indeed, these devices are pervasive in interpersonal communication (DeMeijer, 1989; 1991; Field and Walden, 1982; Saarni, 1982; Scherer, 1981; 1986; Scherer and Oshinsky, 1977).

Children used a greater variety of expressive devices in their sad performances than in their happy performances, a finding that is consistent with children’s greater linguistic elaboration of negative compared to positive events (Dunn and Brown, 1994). Incidentally, children in the present study exhibited more hesitations, complaints and private speech while preparing for their sad renditions than for their happy renditions. Some children seemed to think that singing a song poorly would qualify as a sad rendition: “. . . I shouldn’t sing it well?” (7-year-old boy); “Like not sing it that good?” (8-year-old girl). Similar interpretations were reflected in the use of speech-like melody and suppressed contour, adopting a unique voice quality and manipulating the lyrics (e.g. The ABC Song: “. . . next time don’t sing with me” instead of “. . . next time won’t you sing with me?”). Poor singing as a means of making someone feel sad parallels children’s attempt to do a job well as a means of making someone feel better (Averill and More, 1993; Saarni, 1992).

To children, the act of singing a familiar song may be a happy experience, much like their interpretation of emotionally neutral facial expressions (Reichenbach and Masters, 1983) and stories (Harris, 1993; Reichenbach and Masters, 1983) as happy. Children’s spontaneous comments in the present study gave some suggestion of this perspective: “To make you feel happy, you want me just to sing it normally?” (12-year-old girl). Whether children consider singing as a happy activity because of associations with maternal singing (Trehub and Schellenberg,
1995; Trehub and Trainor, 1998; Trehub et al., 1993) or with celebratory circumstances (e.g. birthdays) is an open question.

Everyday communication as a source of expressive elements sheds light on the use of other expressive devices. For example, 6–7-year-old children were unique in creating new lyrics and adding emotion-evoking words to distinguish their sad from their happy performances. This verbal strategy in singing parallels the domination of semantic over paralinguistic cues in children’s production and comprehension of emotion in speech (Bugental, Kaswan, Love and Fox, 1970; Reilly, 1992; Solomon and Ali, 1972). Portamento, an expressive device that enhances the emotionality of vocal performances in Western art music (Desain and Honing, 1996), is roughly equivalent to a speech glide, although the former requires finer control of melodic modulation from one specific pitch to another (Randel, 1986, p. 649). The shared properties of portamento and common speech glides may have promoted the use of this musical device in children with adequate vocal control (i.e. good singers as young as 4–5 years of age and all 11–12-year-olds).

Some children sang their sad version with an “angry” voice accompanied by an “angry” facial expression, a situation that may stem from confusion between the negative emotions of sadness and anger (Faber and Moely, 1979, cited in Bretherton et al., 1986; Kurdek and Rodgon, 1975). This explanation is at odds with the early appearance of sad and angry expressions in daily interactions (Emde et al., 1978; Fabes, Eisenberg, Nyman and Michalieu, 1991; Harris, 1989; Lewis, 1993) and pretend play (Bretherton et al., 1981) as well as children’s early labelling of “sad” and “angry” feelings in themselves and others (Bretherton et al., 1981; Bretherton and Beeghly, 1982; Fabes et al., 1991; Radke-Yarrow and Zahn-Waxler, 1973, cited in Bretherton et al., 1986). Perhaps confusion arises from circumstances that may lead children to become sad. For example, children commonly react with sadness to their parents’ vocal or facial expressions of anger (see Bretherton et al., 1986; Bretherton et al., 1981). Thus, children may have expected their angry expressions to induce sadness in their adult audience. For similar reasons, children may have expected their negative lyrics to generate sadness in the listener.

Adults use direct expressions of emotion (nasal voice, crying voice) to induce empathic feelings in their communication partners, which is what the majority of children seemed to do in the present study. Perhaps these children understood that emotional displays can evoke the same feeling in others through empathic experience (Harris, 1989; Hoffman, 1981; Lewis, 1993; Ratner and Stettner, 1991; Saarni, 1993). Whether children drew on their own experience of happiness or sadness or used empathic means to induce particular emotions in others, it is clear that they attempted to anticipate the consequences of their emotional expressions. Such role-taking is critical to the successful communication of emotion (Morency and Krauss, 1982).

Increasing age was associated with decreased use of idiosyncratic expressions and greater consensus in the use of expressive symbols. For example, the type of unique voice used by 4–5-year-olds varied from animal sounds and piercing voice quality in their happy renditions to falsetto, squeaky, throaty and mumbling voice in their sad renditions. By 8 years of age, however, unique voice, ranging from angry voice to crying, throaty and mumbling voice, depicted sadness exclusively. This developmental pattern may reflect children’s increasing
awareness of conventional expressions and their role as emotional elicitors, knowledge that is central to the socialisation of emotion (Lewis and Michalson, 1983).

Increased maturity, socialisation and musical knowledge also influenced children’s use of conventional expressive devices in music. For example, the production of legato, that is, “played smoothly with no separation between successive notes” (Randel, 1986, p. 443), requires fine control of vocal muscles and breathing (Bunch, 1995); 4–5-year-olds’ use of non-legato and 11–12-year-olds’ use of legato likely reflect maturational changes in vocal control. Leg movement (e.g. swinging, kicking) while singing, which occurred for 4- and 5-year-olds, likely reflects young children’s tendency to react to music with motion (Cohen, 1980/1981; Metz, 1989). Although expressive body movements (e.g. swaying) were also used by ordinary singers who were 8–10 years old, they were not used by good singers in this age-group, nor by any 11–12-year-old singers. These findings are consistent with the decreased use of movement with increasing musical maturity (Cohen, 1980/1981). The unexpected absence of expressive movement in 6–7-year-olds may have resulted from their inclination to change the lyrics of songs, a strategy that demanded considerable concentration. Perhaps 11–12-year-olds’ tendency to look downward stems from avoidance of eye-gaze with their audience, a behavioural inhibition arising from self-consciousness in early adolescence (Elkind, 1967; Fenigstein, Scheier and Buss, 1975; Hamer and Bruch, 1994).

Gender differences were evident in girls’ greater use of conventional musical devices such as legato and portamento and boys’ greater use of non-musical devices such as frowning, postural change and unique voice. Whether these differences stem from girls’ superior singing skills (girls outnumbering boys among “good” singers), from greater involvement in music (Campbell, 1991; Larsen, 1987; O’Neill and Boulton, 1996), or from their greater conventionality (Brody, 1985; Lott, 1978; Maccoby and Jacklin, 1974) relative to boys is an open question. Regardless of the age and gender differences that were observed, children consistently used tempo, dynamics and pitch to differentiate their happy from their sad performances.

In sum, we identified numerous expressive devices for conveying happiness and sadness in children differing in age, gender and singing ability. Our approach differs from that of other investigators by including facial, gestural, vocal, linguistic and musical devices rather than focusing solely on a single modality (Field and Walden, 1982; Gottman, 1993; Pittam and Scherer, 1993). Moreover, our findings confirm the importance of multimodal approaches to the developmental study of emotion (Lewis, 1993; Pittam and Scherer, 1993). Children’s expression of emotion in the present study was not a reflection of their emotional state. Rather, it was a portrayal designed to induce a particular emotion in others. Such communication between singers and their audience has much in common with the portrayals that have been used to study emotional communication in interpersonal contexts (Cole, 1985; Hochschild, 1983, cited in Oatley, 1993; Saarni, 1992) and in musical performances (Gabrielsson and Juslin, 1996; Juslin, 1997a; Kendall and Carterette, 1990; Ohgushi and Hattori, 1996; Senju and Ohgushi, 1987). Although such portrayals revealed a wide range of expressive devices in songs performed by children 4–12 years of age, their efficacy in conveying the intended emotions remains to be determined.
Authors’ Note

This research was supported by the Social Sciences and Humanities Research Council of Canada. We are grateful to Jane Campbell for assistance in data collection and coding and to Mandy Orchard, Tanya May and Marianne Fallon for assistance in coding. Correspondence about this paper can be addressed to Mayumi Adachi, Faculty of Education and Human Sciences, Yamanashi University, 4-4-37, Takeda, Kofu, Yamanashi 400-8510, Japan (E-mail: adachi@grape.kkb.yamanashi.ac.jp) or to Sandra E. Trehub, Department of Psychology, University of Toronto at Mississauga, Mississauga, Ontario, Canada, L5L 1C6 (E-mail: trehub@erin.utoronto.ca).

References


