Decoding the Expressive Intentions in Children’s Songs

MAYUMI ADACHI & SANDRA E. TREHUB
University of Toronto

Adults and children were exposed to separate visual and auditory cues from paired renditions of familiar songs by young, untrained singers who attempted to express happiness and sadness. Same-age children and adults decoded the expressive intentions of 8- to 10-year-old singers with comparable accuracy (Experiment 1). For performances by 6- to 7-year-old singers, same-age children were less accurate decoders than were adults (Experiment 2). The younger performers also provided poorer cues to the intended emotion than did the older performers. Moreover, 6- to 7-year-olds were less accurate than 8- to 10-year-olds at decoding the performances of 8- to 10-year-old singers (Experiment 3). The findings indicate that, although young children successfully produce and interpret happy and sad versions of familiar songs, 6- to 7-year-old children are less proficient than are 8- to 10-year-old children and adults.

Music Perception
Winter 2000, Vol. 18, No. 2, 213-224
© 2000 by the Regents of the University of California
All rights reserved.

Address correspondence to Mayumi Adachi, Faculty of Education and Human Sciences, Yamanashi University, 4-4-37, Takeda, Kofu, Yamanashi, 400-8510, Japan (adachi@edu.yamanashi.ac.jp) or to Sandra E. Trehub, Department of Psychology, University of Toronto at Mississauga, Mississauga, Ontario, Canada L5L 1C6 (sandra.trehub@utoronto.ca).

ISSN: 0730-7829. Send requests for permission to reprint to Rights and Permissions, University of California Press, 2000 Center St., Ste. 303, Berkeley, CA 94704-1223.
Although much is known about children's understanding of emotional expression (for reviews, see Denham, 1998; Saarni, 1999), most of the research has focused on verbal and facial cues to emotion. The general consensus is that young children are relatively adept at expressing and comprehending basic feeling states such as happiness and sadness. By 4 or 5 years of age, they can even pose a number of emotional expressions (Lewis, Sullivan, & Vasen, 1987). They are much less adept, however, when their own and others' emotional expressions are at odds with inner states, as in attempts to feign happiness after receiving a disappointing gift (Cole, 1986; Davis, 1995). The disappointed recipient in such instances must have adequate control over expressive behaviors as well as knowledge of the impact of particular expressive behaviors on others.

Expressive aspects of music performance have analogous elements of pretense in the sense that performers attempt to project particular feelings whether or not these match their own internal states. Successful communication is achieved when the audience appropriately discerns the performer's expressive intentions. In music, as in other domains, conventions governing emotional expressiveness must be known by performer and audience alike (Gabrielsson & Juslin, 1996; Juslin, 1997a, 1997b; Senju & Ohgushi, 1987). Children, by virtue of their limited experience, might have incomplete knowledge of these expressive conventions.

We know, however, that when young children listen to performances of instrumental and vocal music, they interpret "happy" or "sad" musical meanings in much the same way as adults (Cunningham & Sterling, 1988; Dolgin & Adelson, 1990; Kratus, 1993). Children can also decode happiness and sadness in the expressive body movements of dancers (Boone & Cunningham, 1998), an ability that should facilitate the interpretation of live musical performances. Adachi and Trehub (1998) demonstrated, moreover, that children 4–12 years of age generate highly contrastive renditions of a familiar song when they attempt to convey happiness and sadness to an unfamiliar listener. In such situations, they produce happy renditions that are faster, louder, and at a higher pitch level than are their sad renditions. Children also alter their posture and facial expression across emotional contexts, from upright and smiling during their happy performances to slumping and frowning during their sad performances. Interestingly, children make little use of music-specific devices (e.g., major/minor, staccato/legato), relying instead on expressive devices used in interpersonal communication.

Although the children in Adachi and Trehub's (1998) study generated contrastive performances for their happy and sad renditions, it is unclear whether their performances were effective in portraying the intended emotions. Accordingly, we sought to determine whether the vocal and visual aspects of these performances were interpretable to same-age children and
to adults. On the basis of children’s ability to interpret the expressive body movements of skilled performers (Boone & Cunningham, 1998) and the posed facial expressions of other children (Proft & Whissell, 1991), they were expected to succeed in decoding the visual cues that accompany same-age children’s performances of familiar songs. It is possible that children would be more accurate than adults in this respect, just as they are superior at decoding the facial expressions of same-age children (Proft & Whissell, 1991).

Previous research on children’s interpretation of expressive musical performances involved professional vocalists or instrumentalists (Cunningham & Sterling, 1988; Dolgin & Adelson, 1990; Kratus, 1993) rather than young, untrained vocalists. Nevertheless, children’s use of expressive speech devices in their sung performances (Adachi & Trehub, 1998) and their ability to decode vocal paralanguage (Dimitrovsky, 1964; Morton & Trehub, in press) made it likely that same-age children would successfully distinguish happy from sad renditions of songs.

There are claims that adults more accurately interpret the expressive intentions of expert performers on the basis of visual compared with auditory cues (Davidson, 1993; Ohgushi & Hattori, 1996). On the one hand, the facial expressions and body movements of untrained performers might provide a clearer index of expressive intentions than would vocal cues such as tempo, loudness, and pitch level. On the other hand, unsophisticated child performers might provide exaggerated vocal cues in their performances, as they do when singing to their infant siblings (Trehub, Unyk, & Henderson, 1994). Thus, it was difficult to predict differences in decoding accuracy from visual cues alone compared with auditory cues alone.

A final goal was to compare girls’ and boys’ portrayal and decoding of emotion in sung performances. Effective portrayal could be indexed by accurate decoding of the intended emotion by children and adults. On the basis of girls’ reportedly greater control of emotional expressiveness relative to boys (Cole, 1986; Denham, 1998; Saarni, 1999), we expected more accurate interpretations of girls’ sung performances by child as well as adult viewers and listeners. Despite differences between the sexes in the socialization of emotion (Brody & Hall, 1993), girls seem no better at decoding emotion than are boys (Saarni, 1989; Thompson, 1989). Thus, there was no basis for expecting girls’ superior decoding of emotion.

**Experiment 1**

The child singers from Adachi and Trehub (1998) who were of greatest interest for the purposes of the present experiment were 8- to 10-year-olds who had been classified as “ordinary” as opposed to “good” singers. Child
singers had been designated "good" if they could provide two fully in-tune versions of a familiar song, one with the words and the other without the words (la la la); the remaining children had been designated "ordinary" singers. The listeners and viewers in the present study were same-age children and a comparison group of adults.

METHOD

Participants

The adults were 30 college students (15 men, 15 women) from 18 to 25 years of age (M = 19 years) who received partial course credit for their participation. There were 30 child participants (15 boys, 15 girls) 8–10 years of age (M = 9 years, 4 months) whose families had volunteered to participate in campus research. Music training (private lessons) averaged 1.33 years for adults and 1.07 years for children.

Apparatus

Testing took place in a quiet room. Visual stimuli were presented by means of a video cassette player (Panasonic ST-S1900 N) and a 30-cm color monitor (Panasonic CT-1930 V). Auditory stimuli were presented through digitally referenced headphones (Sony MDR CD550) by means of a Macintosh 8100/100 computer and PsyScope software.

Stimuli

The stimuli consisted of auditory or visual components of the performances of 20 child singers (11 boys, 9 girls) from Adachi and Trehub (1998); nine children were 8 years old, seven were 9 years old, and four were 10 years old (M = 8 years, 10 months). The children had sung "Twinkle, Twinkle, Little Star" or "The ABC Song" (their choice), both of which have the same tune. These children were relatively unskilled singers (deemed "ordinary" singers in Adachi & Trehub, 1998) in the sense that they had failed, in an initial phase, to produce perfectly in-tune versions of a familiar song. Subsequently, they sang two versions of the song (in counterbalanced order), the first to make the experimenter feel happy, and the second to make her feel sad. The happy and sad performances of each child were randomly ordered (with respect to which was first or second) and presented in pairs to the viewing or listening audience. There were two random orders of the videotaped (no sound) pairs. The digital audio pairs were randomly ordered for each listener.

Procedure

All participants were tested individually. Half received the viewing task before the listening task; the other half received the reverse order. Before the viewing task, participants were told that they would see soundless performances of 20 children singing two versions of "Twinkle" or "ABC," and they were to judge which performance appeared to be happier. Before the listening task, participants were told that they would hear 20 performances of children singing two versions of "Twinkle" or "ABC," and they were to judge which performance sounded happier.

RESULTS AND DISCUSSION

A four-way analysis of variance with repeated measures was used to evaluate the effects of presentation medium (auditory or visual), age of
decoder (child or adult), sex of decoder, and sex of singer. There was a main effect of presentation medium, $F(1, 56) = 85.29, p < .0001$, with the proportion of correct judgments being greater for auditory ($M = .83, SD = .07$) than for visual ($M = .73, SD = .09$) materials. There was a main effect of sex of singer, $F(1, 56) = 16.68, p < .0001$, with greater accuracy (proportion of correct judgments) evident for female singers ($M = .83, SD = .09$) than for male singers ($M = .77, SD = .09$), and an interaction between presentation medium and sex of singer, $F(1, 56) = 19.87, p < .0001$ (see Figure 1). Post hoc paired $t$-tests revealed that the advantage for female singers was restricted to visual presentation, $t(59) = -5.44, p < .0001$. There were no other main effects or interactions.

Children and adults succeeded in decoding the expressive intentions of 8- to 10-year-old singers, which indicates that young children can convey happiness and sadness in their renditions of a familiar song. However, girls made more effective use of visual expressive cues in their sung performances than did boys, as reflected in the greater ease of decoding girls’ expressive intentions from visual information alone. The latter finding is consistent with previous evidence of girls’ superior encoding of facial expressions (Zuckerman & Przewuzman, 1979).

Experiment 2

The successful transmission and interpretation of emotion by children 8 to 10 years of age provided the impetus for a comparable evaluation of younger children. Accordingly, we sought to determine whether 6- to 7-year-old singers would be as effective as 8- to 10-year-olds in

Fig. 1. Proportion correct judgments of emotion from auditory and visual cues in the sung performances of 8- to 10-year-old children. Error bars are standard errors.
communicating their emotional intentions to same-age and adult audiences.

**METHOD**

**Participants**

Adults and children were recruited in the same manner as Experiment 1. There were 49 adults (17 male, 32 female) 18–42 years of age (M = 23 years) and 48 children (24 boys, 24 girls) 6 years, 6 months to 7 years, 11 months of age (M = 7 years). Music training averaged 2.8 years for adults and 0.4 years for children.

**Apparatus and Stimuli**

The apparatus was identical to that of Experiment 1 except for the use of smaller headphones (Koss WM/60) with the younger children. The stimuli consisted of auditory and visual components of the performances of 20 6- to 7-year-old “ordinary” singers from Adachi and Trehub (1998). These children had been given the choice of singing “Twinkle, Twinkle, Little Star,” “The ABC Song,” or “Baa Baa Black Sheep,” all of which have the same tune. Unlike Experiment 1, in which participants completed auditory and visual decoding tasks, participants in the present experiment completed either the auditory decoding task (25 adults, 24 children) or the visual decoding task (24 adults, 24 children). In all other respects, the procedure was identical to that of Experiment 1.

**RESULTS AND DISCUSSION**

The auditory decoding score of one 6-year-old boy was excluded because of its extreme value. A four-way analysis of variance with repeated measures—presentation medium, sex of decoder, and age of decoder as between-subjects factors and sex of singer as a within-subject factor—revealed a main effect of presentation medium, F(1, 88) = 16.33, p < .0001, with greater accuracy (proportion correct) on auditory (M = .77, SD = .16) than on visual (M = .66, SD = .10) materials, in line with the findings of Experiment 1. There was a main effect of age of decoder, F(1, 88) = 7.74, p = .007, with adults performing more accurately (M = .76, SD = .11) than children (M = .67, SD = .16), unlike the situation in Experiment 1. There was also an effect of sex of decoder, F(1, 88) = 8.19, p = .005, with female decoders performing better (M = .75, SD = .13) than male decoders (M = .67, SD = .15). There was a significant two-way interaction between presentation medium and sex of singer, F(1, 88) = 9.75, p = .002 (see Figure 2). Post hoc paired t-tests revealed superior decoding of girls’ expressive auditory cues (M = .79, SD = .17) relative to those of boys (M = .74, SD = .17), t(47) = 2.36, p = .022, and superior decoding of boys’ expressive visual cues (M = .69, SD = .14) relative to those of girls (M = .63, SD = .13), t(47) = 2.33, p = .024.

In short, children 6–7 years of age can use auditory and visual expressive devices to transmit happiness and sadness in their performances of familiar
songs. However, the superior decoding of boys’ visual cues relative to those of girls is at odds with the findings of Experiment 1, precluding unequivocal interpretation of decoding differences between the sexes in either experiment. Similarly, superior decoding by female relative to male audiences also differs from the findings of Experiment 1. Perhaps female decoding superiority becomes evident only when the task is made difficult by the availability of limited cues.

Adults’ greater success in decoding the expressive intentions of 6- to 7-year-olds relative to that of same-age children may be attributable, in part, to young singers’ less effective use of expressive cues relative to the older singers in Experiment 1. The performance of adult listeners (9 men, 16 women) in the present experiment was compared with that of adults who had received the audio task first in Experiment 1 (7 men, 8 women). Similarly, the performance of adult viewers in the present experiment (8 men, 16 women) was compared with that of adults who had received the video task first in Experiment 1 (8 men, 7 women). A three-way analysis of variance (with sex of singer as a within-subject variable and age of singer and sex of audience as between-subjects variables) revealed no differences in decoding the auditory cues from older and younger children. With respect to the decoding of visual cues, however, there was a significant effect of sex of singer, $F(1, 37) = 4.43, p = .042$, and a two-way interaction between age of singer and sex of singer, $F(1, 37) = 12.05, p = .001$. Post hoc $t$-tests revealed that adults were significantly more accurate on the performances of older girls ($M = .81, SD = .12$) than on the performances of younger girls ($M = .67, SD = .13$), $t(31.78) = -3.47, p = .002$; they were no more accurate,
however, on the performances of older boys ($M = .67, SD = .13$) than on the performances of younger boys ($M = .70, SD = .12$). There are indications, then, of developmental differences in the efficacy of girls' visual cues to emotion.

The provision of less interpretable cues by 6- to 7-year-olds than by 8- to 10-year-olds, as reflected in adults’ decoding accuracy, may have depressed the decoding performance of the 6- to 7-year-olds. It is possible, however, that younger children make less effective use of comparable expressive cues than do older children, an issue that could not be resolved from the present data set. Accordingly, we evaluated 6- to 7-year-olds’ decoding of expressive cues from 8- to 10-year-old children in a subsequent experiment.

Experiment 3

The principal purpose of the present experiment was to ascertain whether 6- to 7-year-old children are less effective decoders than are 8- to 10-year-olds. Accordingly, 6- to 7-year-olds were evaluated on the auditory decoding task of Experiment 1, which involved the sung performances of 8- to 10-year-old children.

METHOD

Participants

The participants were 24 children (12 boys, 12 girls) who were 6 years, 7 months to 7 years, 10 months of age ($M = 7$ years, 2 months). With the exception of a single child who had taken music lessons for 1 year, the children had not experienced formal musical training. A further three children were excluded from the data set because of their inattentiveness throughout the test session.

Stimuli

The auditory materials from Experiment 1, which were derived from the performances of 8- to 10-year-old “ordinary” singers, served as stimuli.

Procedure

The procedure was identical to that used for the auditory decoding tasks in Experiments 1 and 2.

RESULTS AND DISCUSSION

Data from 6- to 7-year-olds were examined together with previous data obtained from 8- to 10-year-olds and adults on the same auditory decoding task (Experiment 1). A three-way analysis of variance with repeated measures, including age and sex of decoder as between-subjects variables and
sex of singer as a within-subject variable, revealed a significant effect of age of decoder, F(2, 48) = 6.24, p = .004. Bonferroni tests (α = .05) revealed poorer decoding accuracy among 6- to 7-year-olds (M = .75, SD = .17) than among 8- to 10-year-olds (M = .84, SD = .07) and adults (M = .83, SD = .07), who did not differ from one another (see Figure 3). No other main effects or interactions were significant. It is clear, then, that 6- to 7-year-olds make less effective use of auditory cues when decoding the intended emotions in children’s informal songs.

General Discussion

The present findings indicate that young children’s expressive renditions of familiar songs are interpretable to same-age children and to adults. Not only do children use distinctive vocal and visual devices when producing “happy” and “sad” versions of familiar songs (Adachi & Trehub, 1998); the resulting performances are readily decoded by 6- to 10-year-old children and adults. Obviously, children have the requisite skill for portraying emotion, vocally and visually, to child and adult audiences. It is clear, moreover, that young children’s ability to discern musical performers’ expressive intentions does not depend on professional performances such as those used in previous research (Cunningham & Sterling, 1988; Dolgin & Adelson, 1990; Kratus, 1993). Rather, the sung performances of untrained children are sufficient.

Adults’ and children’s decoding accuracy was greater for auditory cues than for visual cues from the performances of 8- to 10-year-olds (Experi-
ment 1) and 6- to 7-year-olds (Experiment 2). This finding contrasts with the counterintuitive claim that adults achieve greater accuracy when decoding visual rather than auditory expressive cues from expert musical performances (Davidson, 1993; Ohgushi & Hattori, 1996). No doubt the performer’s repertoire of visual and vocal expressive devices, the nature of the musical materials, and the knowledge of the audience contribute jointly to the relative efficacy of auditory and visual cues.

Adults’ greater decoding accuracy on the performances of older relative to younger children implies that 6- to 7-year-old singers provide less differentiated cues to their happy and sad intentions than do 8- to 10-year-old children. Although adults’ decoding accuracy did not exceed that of 8- to 10-year-old children on performances from that age group (Experiment 1), adults’ accuracy did exceed that of 6- to 7-year-old children on performances from the latter age group (Experiment 2). Comparisons of all three age groups on the auditory materials from 8- to 10-year-olds revealed that 6- to 7-year-olds were less effective decoders than were older children and adults (Experiment 3). These findings concur with Dolgin and Adelson’s (1990) report of poorer decoding of emotional meanings in music by 7-year-olds than by 9-year-olds. It is clear, then, that 6- to 7-year-olds have decoding as well as encoding limitations in relation to expressive performances of familiar songs. No doubt, age-related differences would be greater in open-ended tasks (e.g., naming the intended emotion) than in forced-choice tasks such as those in the present study (i.e., selecting the happier of two performances).

With respect to differences between the sexes, the situation was less clear. Overall, there was no indication that girls were more effective at portraying or decoding emotion than were boys. There were suggestions, however, that 8- to 10-year-old girls provided clearer visual expressive cues than did boys (Experiment 1), which is in line with girls’ reportedly greater facility in expressing emotion (Brody & Hall, 1993; Cole, 1986; Denham, 1998). Although the performances of younger girls and boys led to comparable levels of decoding accuracy, the girls seemed to provide more differentiated auditory cues and the boys more differentiated visual cues (Experiment 2). Finally, when the available expressive cues were less than optimal, as was the case for the performances of younger children, female audiences seemed to have an advantage over male audiences, a finding that is consistent with greater female sensitivity to emotional expression (Brody & Hall, 1993).

In short, the present findings confirmed that children’s happy and sad renditions of familiar songs are interpretable by same-age children and by adults on the basis of visual or vocal expressive cues. Children 8 and older seem to generate more interpretable performances than do those 7 and younger. With respect to performances such as these, decoding accuracy improves with age, reaching adult levels by 8 to 10 years. It is unclear,
however, whether these age differences would be maintained if visual and vocal cues were made available to prospective decoders. For expert performances, moreover, adults’ greater implicit, if not explicit, knowledge of musical performance conventions would likely lead to greater decoding discrepancies between children and adults.¹

References


¹ Funding for this research was provided by the Social Sciences and Humanities Research Council of Canada to Sandra E. Trehub. Portions of this research were presented at the meeting of the Japanese Society for Music Perception and Cognition in November, 1998. We are grateful to Leanne Kahro for assistance in the testing of children and to Marianne Fallon for assistance in computer programming.


