A peer-reviewed version of this preprint was published in PeerJ on 2 April 2020.

View the peer-reviewed version (peerj.com/articles/8773), which is the preferred citable publication unless you specifically need to cite this preprint.

Nagels L, Gaudrain E, Vickers D, Matos Lopes M, Hendriks P, Başkent D. 2020. Development of vocal emotion recognition in school-age children: The EmoHI test for hearing-impaired populations. PeerJ 8:e8773 https://doi.org/10.7717/peerj.8773

Vocal emotion recognition in school-age children: normative data for the EmoHI test

Leanne Nagels[1,2], Etienne Gaudrain[3,2], Debi Vickers[4], Marta Matos Lopes[5,6], Petra Hendriks[1], and Deniz Başkent[2]

[1]**Center for Language and Cognition Groningen (CLCG), University of Groningen, Groningen, The Netherlands**
[2]**Department of Otorhinolaryngology/Head and Neck Surgery, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands**
[3]**CNRS, Lyon Neuroscience Research Center, Université de Lyon, Lyon, France**
[4]**Clinical Neurosciences Department, University of Cambridge, Cambridge, UK**
[5]**Hearbase Ltd, Hearing specialists, Kent, UK**
[6]**The Ear Institute, University College London, London, UK**

Corresponding author:
Leanne Nagels[1,2]

Email address: leanne.nagels@rug.nl

ABSTRACT

Traditionally, emotion recognition research has primarily used pictures and videos while audio test materials have received less attention and are not always readily available. Particularly for testing vocal emotion recognition in hearing-impaired listeners, the audio quality of assessment materials may be crucial. Here, we present a vocal emotion recognition test with non-language specific pseudospeech productions of multiple speakers expressing three core emotions (happy, angry, and sad): the EmoHI test. Recorded with high sound quality, the test is suitable to use with populations of children and adults with normal or impaired hearing, and across different languages. In the present study, we obtained normative data for vocal emotion recognition development in normal-hearing school-age (4-12 years) children using the EmoHI test. In addition, we tested Dutch and English children to investigate cross-language effects. Our results show that children's emotion recognition accuracy scores improved significantly with age from the youngest group tested on (mean accuracy 4-6 years: 48.9%), but children's performance did not reach adult-like values (mean accuracy adults: 94.1%) even for the oldest age group tested (mean accuracy 10-12 years: 81.1%). Furthermore, the effect of age on children's development did not differ across languages. The strong but slow development in children's ability to recognize vocal emotions emphasizes the role of auditory experience in forming robust representations of vocal emotions. The wide range of age-related performances that are captured and the lack of significant differences across the tested languages affirm the usability and versatility of the EmoHI test.

INTRODUCTION

Children's development of emotion recognition has been studied extensively using visual stimuli, such as pictures or sketches of facial expressions, or audiovisual materials (e.g., Nowicki and Duke, 1994), and particularly with clinical groups, such as autistic children (e.g., Harms et al., 2010). However, not much is known about the development of vocal emotion recognition (Scherer, 1986). Children have been reported to reliably recognize vocal emotions already from the age of 5 years on, but this ability continues to develop to adult-like levels throughout childhood (Tonks et al., 2007; Sauter et al., 2013). Based on earlier research on the development of voice perception (Mann et al., 1979; Nittrouer et al., 1993), children's performance may be lower compared to adults due to differences in their weighting of acoustic cues and a lack of robust representations of auditory categories. For instance, Morton and Trehub (2001) showed that, when acoustic cues and linguistic content contradict the emotion they convey, children mostly rely on linguistic content to judge emotions, whereas adults mostly rely on affective prosody. In addition, children and adults both perform better in facial than vocal emotion recognition tasks (Nowicki and Duke,

1994). All of these observations combined indicate that the formation of robust representations for vocal emotions is highly complex and possibly a long-lasting process even in typically developing children.

Research with hearing-impaired children has shown that they do not perform as well on vocal emotion recognition compared to their normal-hearing peers (Dyck et al., 2004; Hopyan-Misakyan et al., 2009; Nakata et al., 2012; Chatterjee et al., 2015). Hopyan-Misakyan et al. (2009) showed that children with cochlear implants (CIs) performed as well as their normal-hearing peers on facial emotion recognition but scored significantly lower on vocal emotion recognition. Facial emotion recognition seems to generally develop faster than vocal emotion recognition (Nowicki and Duke, 1994), particularly in hearing-impaired children (Hopyan-Misakyan et al., 2009), which may indicate that visual emotion cues are perceptually more prominent or easier to categorize than vocal emotion cues. A higher reliance on visual emotion cues as compensation for degraded auditory input, as emotion recognition in daily life is usually multimodal for which visual emotion cues can often be sufficient, may lead to less robust representations of vocal emotions. Furthermore, Nakata et al. (2012) found that children with CIs had difficulties primarily with differentiating happy from angry vocal emotions. This difference may be related to a higher reliance on differences in speaking rate to categorize vocal emotions, as this cue differentiates sad from happy and angry vocal emotions but is similar for the latter two emotions. Therefore, hearing loss also seems to influence the weighting of different acoustic cues, and hence likely also affects the formation of representations of vocal emotions.

As most research on the development of emotion recognition has used visual materials such as pictures or videos, good-quality audio materials are scarce. For normal-hearing listeners, the audio quality may only have a small effect on performance, but for testing hearing-impaired populations it may be highly important. Hence, we recorded high sound quality vocal emotion recognition test stimuli produced by multiple speakers with three basic emotions (happy, angry, and sad) that are suitable to use with hearing-impaired children and adults: the EmoHI test. We aimed to investigate how school-age children's ability to recognize vocal emotions develops with age and to obtain normative data for the EmoHI test for future applications, for instance, with clinical populations. In addition, we tested children of two different native languages, namely Dutch and English, to investigate potential cross-language effects.

METHODS

Participants

Fifty-eight Dutch children and 25 English children between the ages of 4 to 12 years, and 15 Dutch adults and 15 English adults participated in the study. All participants were monolingual speakers of Dutch or English and reported no hearing or language disorders. Normal hearing (hearing thresholds at 20 dB HL) was screened with pure-tone audiometry at octave-frequencies between 500 and 4000 Hz. The study was approved by local ethics committees of the participating institutions. A written informed consent form was signed by the parents of children and adult participants before data collection.

Stimuli and Apparatus

We made recordings of six native Dutch speakers producing two non-language specific pseudospeech sentences using three core emotions (happy, sad, and angry), and a neutral emotion (not used in the current study). All speakers were native monolingual speakers of Dutch without any discernable accent and did not have any speech, language, or hearing disorders. Speakers gave written informed consent for the distribution and sharing of the recorded materials. To keep our stimuli relevant to emotion perception literature, the pseudospeech sentences that we used, *Koun se mina lod belam* [kʌun sə minaː lɔt beːlɑm] and *Nekal ibam soud molen* [neːkɑl ibɑm sʌut moːlɛn], were taken from the Geneva Multimodal Emotion Portrayal (GEMEP) Corpus by Bänziger and Scherer (2010). Speakers were instructed to produce the sentences in a happy, sad, angry, or neutral manner using emotional scripts that were also used for the GEMEP corpus stimuli (Scherer and Bänziger, 2010). The stimuli were recorded in an anechoic room at a sampling rate of 44.1 kHz. We selected the productions which received the highest accuracy scores of the four highest-rated speakers based on an online survey with Dutch and English adults. Table 1 shows an overview of these four selected speakers' demographic information and voice characteristics. The neutral productions and the productions of the other two speakers were part of the online survey, and are available with the stimulus set, but were not used in the current study to simplify the task for children. Our final set of stimuli consisted of 36 experimental stimuli with three items (combinations of two times

one sentence and one time the other sentence) per emotion and per speaker (3 items x 3 emotions x 4 speakers) and 4 practice stimuli with one item per speaker that were used for the training session.

Speaker	Age	Gender	Height	Average F0	F0 range
T2	36	F	1.68 m	302.23 Hz	200.71 - 437.38 Hz
T3	27	M	1.85 m	166.92 Hz	100.99 - 296.47 Hz
T5	25	F	1.63 m	282.89 Hz	199.49 - 429.38 Hz
T6	24	M	1.75 m	167.76 Hz	87.46 - 285.79 Hz

Table 1. Overview of the speakers' demographic information and voice characteristics.

Procedure

Children were tested in a quiet room at their home, and adults were tested in a quiet testing room at the two universities. The present experiment is part of a larger project (PICKA) on voice and speech perception conducted by the UMCG for which data were collected from the same population of children and adults in multiple experiments (Nagels et al., in review). The experiment started with a training session consisting of 4 practice stimuli and was followed by the test session consisting of 36 experimental stimuli. The total duration of the experiment was approximately 6 to 8 minutes. All items were presented to participants in a randomized order.

The experiment was conducted on a laptop with a touchscreen using a child-friendly interface that was developed in Matlab (Figure 1). The auditory stimuli were presented via Sennheiser HD 380 Pro headphones and calibrated to a sound level of 65 dBA. In each trial, participants heard a stimulus and then had to indicate which emotion was conveyed by clicking on one of three corresponding clowns on the screen. Visual feedback on the accuracy of responses was provided to motivate participants. Participants saw confetti falling down the screen after a correct response, and the parrot shaking its head after an incorrect response. After every two trials, one of the clowns in the back went one step up the ladder until the experiment was finished to keep children engaged and to give an indication of the progress of the experiment.

Figure 1. The experimental interface of the EmoHI test.

Data analysis

Children's accuracy scores were analyzed using the lme4 package (version 1.1.21, Bates et al., 2014) in R. A mixed effects logistic regression model with a three-way interaction between *language* (Dutch and English), *emotion* (happy, angry, and sad), and *age* in decimal years, and random intercepts per participant and per item was computed to determine the effects of language, emotion, and age on children's ability to recognize vocal emotions. We used backward stepwise selection with ANOVA Chi-Square tests to select the best fitting model, starting with the full factorial model, in lme4 syntax: $accuracy \sim language \cdot emotion \cdot age + (1|participant) + (1|item)$, and deleting one fixed factor at a time based on its significance. In addition, we performed Dunnett's tests on the Dutch and the English data with *accuracy* as an outcome variable and *age group* as a predictor variable using the DescTools package (version 0.99.25, Signorell et al., 2016) to investigate at what age Dutch and English children showed adult-like performance.

RESULTS AND DISCUSSION

Model comparison showed that the full model with random intercepts per participant and per item was significantly better than the full model with only random intercepts per participant [$\chi^2(1) = 393$, $p < 0.001$] or only random intercepts per item [$\chi^2(1) = 51.9$, $p < 0.001$]. Backward stepwise selection showed that the best fitting and most parsimonious model was the model with only a fixed effect of *age*, in lme4 syntax: $accuracy \sim age + (1|participant) + (1|item)$. This model did not significantly differ from the full model [$\chi^2(10) = 12.90$, $p = 0.23$] or any of the other models while being the most parsimonious. Figure 2 shows the data of individual participants and the median accuracy scores per age group for the Dutch and English participants. Children's ability to correctly recognize vocal emotions increased as a function of age [z-value = 8.91, estimate = 0.30, SE = 0.034, $p < 0.001$]. We did not find any significant effects of language or emotion on children's accuracy scores. Finally, the results of the Dunnett's tests showed that the accuracy scores of Dutch children of all tested age groups differed from Dutch adults [4-6 years difference = -0.47, $p < 0.001$; 6-8 years difference = -0.31, $p < 0.001$; 8-10 years difference = -0.19, $p < 0.001$; 10-12 years difference = -0.15, $p < 0.001$], and the accuracy scores of English children of all tested age groups differed from English adults [4-6 years difference = -0.43, $p < 0.001$; 6-8 years difference = -0.27, $p < 0.001$; 8-10 years difference = -0.20, $p < 0.001$; 10-12 years difference = -0.12, $p < 0.01$].

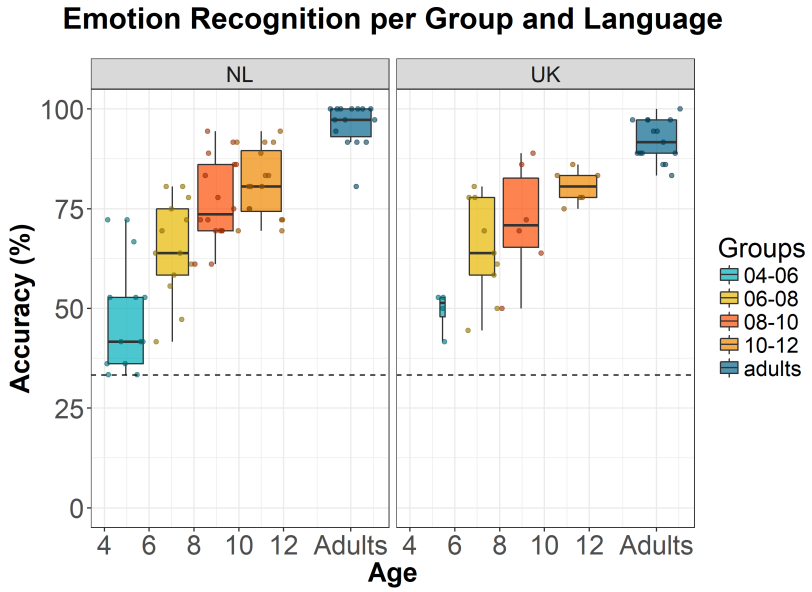

Figure 2. Accuracy scores of participants for emotion recognition per age group and per language (Dutch in the left panel; English in the right panel). The dots show individual data points at participants' decimal age (Netherlands (NL): $N_{children} = 58$, $N_{adults} = 15$; United Kingdom (UK) : $N_{children} = 25$, $N_{adults} = 15$). The boxplots show the median per age group, and the lower and upper quartiles. The whiskers indicate the lowest and highest data points within plus or minus 1.5 times the interquartile range.

Age effect

As shown by our results and the data displayed in Figure 2, children's ability to recognize vocal emotions improved gradually as a function of age. In addition, we found that, on average, even the oldest age group of 10- to 12-year-old Dutch and English children did not show adult-like performance yet. The 4-year-old children that were tested performed at or above chance level while adults generally showed near ceiling level performance, indicating that our test covers a wide range of age-related performances. Our results are in line with previous findings that children's ability to recognize vocal emotions improves as a function of age (Tonks et al., 2007; Sauter et al., 2013). It may be that children require more auditory experience to form robust representations of vocal emotions or rely on different acoustic cues than adults, as was shown for the development of sensitivity to voice cues (Mann et al., 1979; Nittrouer et al., 1993).

It is possible that the visual feedback caused some learning effects, although the correct response was not shown after an error, and learning would pose relatively high demands on auditory working memory, as there were only three items per speaker and per emotion presented in a randomized order.

Language effect

We did not find any cross-language effects between Dutch and English children's development of vocal emotion recognition, even though the materials were produced by Dutch native speakers. Earlier research has demonstrated that although adults are able to recognize vocal emotions across languages, there still seems to be a native language benefit (Van Bezooijen et al., 1983; Scherer et al., 2001). Listeners were better at recognizing vocal emotions that were produced by speakers of their native language than another language. However, these studies used five (Scherer et al., 2001) and nine (Van Bezooijen et al., 1983) different emotions which is likely considerably more complex then differentiating three basic emotions. In addition, the lack of a native language benefit may also be due to the fact that Dutch and English are closely related languages. We are currently collecting data from Turkish children and adults to investigate whether there are any detectable cross-language effects for typologically and phonologically more distinct languages.

Future directions

The results of the current study provide a baseline for the development of vocal emotion recognition for normal-hearing typically developing school-age children using the EmoHI test. Our results show that there is a large but relatively slow development in children's ability to recognize vocal emotions which also brings up the question on which specific acoustic cues children are basing their decisions and how this differs from adults. Future research using machine-learning approaches may be able to further explore such aspects. We are currently collecting data from children with CIs for whom the amount of auditory exposure is reduced due to degraded auditory input. The reduction of auditory exposure may delay or even limit the development of vocal emotion recognition in children with CIs, as some acoustic cues may not be available to hearing-impaired children due to degraded auditory input (Nakata et al., 2012). To conclude, the evident development in children's performance as a function of age and the generalizibility across the tested languages show the EmoHI Tests' suitability for future applications with hearing-impaired or other clinical populations of children and adults across different languages.

ACKNOWLEDGMENTS

We are grateful to all children, parents, and students that participated in the study, the speakers of our stimuli, and Basisschool de Brink in Ottersum, Basisschool de Petteflet, and BSO Huis de B in Groningen for their help with recruiting child participants. We would also like to thank Iris van Bommel, Evelien Birza, Paolo Toffanin, Jacqueline Libert, Jemima Phillpot, and Jop Luberti (illustrations) for their contribution to the development of the game interfaces, and Monita Chatterjee for her advice on recording the sound stimuli. This work was funded by the Center for Language Cognition Groningen (CLCG), a VICI Grant from the Netherlands Organization for Scientific Research (NWO) and the Netherlands Organization for Health Research and Development (ZonMw) (Grant No. 918-17-603), the Medical Research Council (Senior Fellowship Grant S002537/1), and framework of the LabEx CeLyA ("Centre Lyonnais d'Acoustique", ANR-10-LABX-0060/ANR-11-IDEX-0007), and the French National Research Agency.

REFERENCES

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., Dai, B., and Grothendieck, G. (2014). Package 'lme4'. *R Foundation for Statistical Computing, Vienna*, 12.

Bänziger, T. and Scherer, K. R. (2010). Introducing the Geneva Multimodal Emotion Portrayal (GEMEP) Corpus. In *A Blueprint for Affective Computing: A Sourcebook and Manual*, pages 271–294.

Chatterjee, M., Zion, D. J., Deroche, M. L., Burianek, B. A., Limb, C. J., Goren, A. P., Kulkarni, A. M., and Christensen, J. A. (2015). Voice emotion recognition by cochlear-implanted children and their normally-hearing peers. *Hearing research*, 322:151–162.

Dyck, M. J., Farrugia, C., Shochet, I. M., and Holmes-Brown, M. (2004). Emotion recognition/understanding ability in hearing or vision-impaired children: do sounds, sights, or words make the difference? *Journal of Child Psychology and Psychiatry*, 45(4):789–800.

Harms, M. B., Martin, A., and Wallace, G. L. (2010). Facial Emotion Recognition in Autism Spectrum Disorders: A Review of Behavioral and Neuroimaging Studies. *Neuropsychology Review*, 20(3):290–322.

Hopyan-Misakyan, T. M., Gordon, K. A., Dennis, M., and Papsin, B. C. (2009). Recognition of Affective Speech Prosody and Facial Affect in Deaf Children with Unilateral Right Cochlear Implants. *Child Neuropsychology*, 15(2):136–146.

Mann, V. A., Diamond, R., and Carey, S. (1979). Development of voice recognition: Parallels with face recognition. *Journal of Experimental Child Psychology*, 27(1):153–165.

Morton, J. B. and Trehub, S. E. (2001). Children's Understanding of Emotion in Speech. *Child Development*, 72(3):834–843.

Nagels, L., Gaudrain, E., Vickers, D., Hendriks, P., and Başkent, D. (in review). School-age children's development in sensitivity to voice gender cues is asymmetric.

Nakata, T., Trehub, S. E., and Kanda, Y. (2012). Effect of cochlear implants on children's perception and production of speech prosody. *The Journal of the Acoustical Society of America*, 131(2):1307–1314.

Nittrouer, S., Manning, C., and Meyer, G. (1993). The perceptual weighting of acoustic cues changes with linguistic experience. *The Journal of the Acoustical Society of America*, 94(3):1865–1865.

Nowicki, S. and Duke, M. P. (1994). Individual differences in the nonverbal communication of affect: The diagnostic analysis of nonverbal accuracy scale. *Journal of Nonverbal Behavior*, 18(1):9–35.

Sauter, D. A., Panattoni, C., and Happé, F. (2013). Children's recognition of emotions from vocal cues. *British Journal of Developmental Psychology*, 31(1):97–113.

Scherer, K. R. (1986). Vocal affect expression: A review and a model for future research. *Psychological Bulletin*, 99(2):143–165.

Scherer, K. R., Banse, R., and Wallbott, H. G. (2001). Emotion Inferences from Vocal Expression Correlate Across Languages and Cultures. *Journal of Cross-Cultural Psychology*, 32(1):76–92.

Scherer, K. R. and Bänziger, T. (2010). On the use of actor portrayals in research on emotional expression. In *A Blueprint for Affective Computing: A Sourcebook and Manual*, pages 271–294.

Signorell, A., Aho, K., Alfons, A., Anderegg, N., and Aragon, T. (2016). DescTools: Tools for descriptive statistics. R package version 0.99. 18. *R Found. Stat. Comput., Vienna, Austria*.

Tonks, J., Williams, W. H., Frampton, I., Yates, P., and Slater, A. (2007). Assessing emotion recognition in 9–15-years olds: Preliminary analysis of abilities in reading emotion from faces, voices and eyes. *Brain Injury*, 21(6):623–629.

Van Bezooijen, R., Otto, S. A., and Heenan, T. A. (1983). Recognition of Vocal Expressions of Emotion: A Three-Nation Study to Identify Universal Characteristics. *Journal of Cross-Cultural Psychology*, 14(4):387–406.