Do Emotional Voices Move Us? Investigating Bodily Approach-Avoidance Responses

Junko Yakushiji, Hosei University, Japan
Yayoi Watanabe, Hosei University, Japan

The Asian Conference on Psychology & the Behavioral Sciences 2026
Official Conference Proceedings

Abstract

Previous studies on emotional understanding from vocal cues have mainly relied on explicit methods such as using emotion-related words or facial illustrations to identify emotions. However, these approaches may not adequately capture the developmental characteristics of early childhood emotional understanding. Several implicit methods have been proposed to address these limitations. For example, Hiraoka et al. (2019) analyzed center-of-pressure (COP) shifts and found that mothers quickly approached unpleasant and urgent sounds in infants. However, it remains unclear whether similar approach-avoidance responses occur in adults, and whether COP shifts can be used to assess reactions to emotional vocal stimuli. This study aimed to examine whether patterns of approach and avoidance responses vary depending on the type of emotion expressed in vocal stimuli. Twenty-four university students listened to eight types of emotional voices. These vocal stimuli were produced by one male and one female actor uttering “waa” with different emotions (joy, sadness, anger and neutral). The interaction between each vocal stimulus and time was significant. Sadness (female) exhibits a movement that gradually approaches, whereas anger (female) and neutrality (female) exhibit a movement that gradually moves in the avoidance direction. However, the multiple comparison results were not significant and no significant differences were observed for the other voices. Two possible explanations are considered: (1) individual differences, such as sensitivity to emotions or voices and (2) a possible mismatch between the intended emotions in the stimuli and the participants’ perceptions of those emotions.

Keywords: emotion, voice, emotional understanding

iafor
The International Academic Forum
www.iafor.org

Introduction

How do people understand the emotions of others through their voices? Emotional understanding is a crucial component of emotional literacy. Although there is abundant research on understanding emotions through facial expressions, studies focusing on voices remain scarce. Adegawa and Watanabe (2021) investigated the development of emotional understanding when listening to vocal stimuli of basic emotions (joy, sadness, and anger) using emotional vocabulary or facial expressions indicating emotions, targeting elementary school children. The results showed that the emotional understanding of vocal stimuli generally improved in accuracy from grades 1 and 2 to grades 5 and 6. Notably, for joyful voices, the accuracy rate exceeded 90%, even in Grade 1. Watanabe et al. (2024) investigated infant children's abilities to recognize emotions from the voices. The study revealed that by the age of four, over 80% of the children could accurately identify joyful voices, but their ability to recognize negative voices remained underdeveloped.

In developmental studies on emotional understanding from voice, explicit methods have traditionally been applied to evaluate whether individuals can accurately identify emotions conveyed by voice using emotional words or facial expressions. However, there is also the perspective that emotional understanding may already be occurring at a bodily level prior to conscious awareness, leading to proposals for examining emotional understanding using implicit methods (Ikeda & Haryu, 2016). Hiraoka et al. (2019) analyzed center-of-pressure (COP) shifts in mothers when listening to infant vocalizations and identified rapid approach responses to unpleasant and urgent voices. However, it remains unclear whether similar approach/avoidance responses generally occur in adults.

Therefore, this study explored the possibility of measuring emotional understanding in the general adult population using COP shifts for voices containing multiple emotions. We examined whether patterns of bodily approach and avoidance responses varied depending on the type of emotion expressed in the vocal stimuli.

Method

Participants

Twenty-four university students participated in this experiment (9 males and 15 females; average age = 18.83 years [SD = 1.37, Range 18–23]).

Materials

Vocal stimuli were produced by one male and one female actor uttering “waa,” “aa,” and “ee” with four different emotions (joy, sadness, anger, and neutral). These actors had over five years of experience. They expressed their emotional voices while visualizing scenes that aroused emotions. These scenes were created by modifying expressions originally intended for children based on the stories of Sawada et al. (2024) and Hamana and Haryu (2015) to suit adults. The vocal stimuli were recorded using a linear PCM recorder (SONY PCM-A10).

The created vocal stimuli were evaluated by two experts (excluding the author) who selected the most appropriate option from five choices (joy, sadness, anger, neutral, or other). Only those stimuli for which the opinions of both the experts and the author matched

were used. Eight types of vocal stimuli, including two sexes and four emotions, were used in this study. However, as the experts could not agree on the evaluation of the female speaker's neutral voice, the stimuli were reselected and re-evaluated, and only those stimuli that achieved consensus were adopted.

The duration of the vocal stimuli ranged from a minimum of 0.731 s to a maximum of 1.156 s, and their intensity was adjusted to between 60 and 70 dB.

Methods

The vocal stimuli presentation comprised two blocks: Block 1 presented only the male speaker's voice and Block 2 presented only the female speaker's voice, with an interval of approximately 1 min between them. Each block lasted approximately 3 min, during which the participants listened to the vocal stimuli while standing on a GRAVICORDER GW-5000 (Anima). This listening duration was considered the maximum time that participants could comfortably continue listening while standing. The order of presentation for Blocks 1 and 2 is counterbalanced.

Each vocal stimulus was presented four times within each block, beginning with joy and neutral, ending with neutral, followed by joy. This was done to alleviate the psychological burden on the participants and minimize the impact on their subsequent emotional states. The order of presentation within each block was adjusted to prevent bias in the responses to subsequent stimuli based on the emotional category of the preceding stimulus. Each vocal stimulus interval lasted 9 s, during which a cross was displayed.

The experiment was conducted in a soundproof room with participants standing while listening to their voices. The laptop was positioned approximately 130 cm above and 40 cm in front of the GW-5000, and the vocal stimuli were presented on the laptop. The participants were instructed to stand still for 3-min duration during the vocal presentation, moving only as necessary.

Data Analysis

To examine whether approach-avoidance reaction patterns differed according to the type of vocal stimulus, we conducted a repeated-measures analysis of variance. The dependent variable was based on changes in the COP during the 3 s following the sound onset. Specifically, the COP was divided into equal 500 ms time intervals, and the averaged COPs were calculated and applied. The independent variables comprised two factors: vocal stimulus and time. There were eight types of vocal stimuli, with four emotional types (joy, sadness, anger, and neutral) for each speaker's sex (male or female). Time ranged from Bin1 to Bin6. Analysis was performed using JASP 0.95.4.

Approval

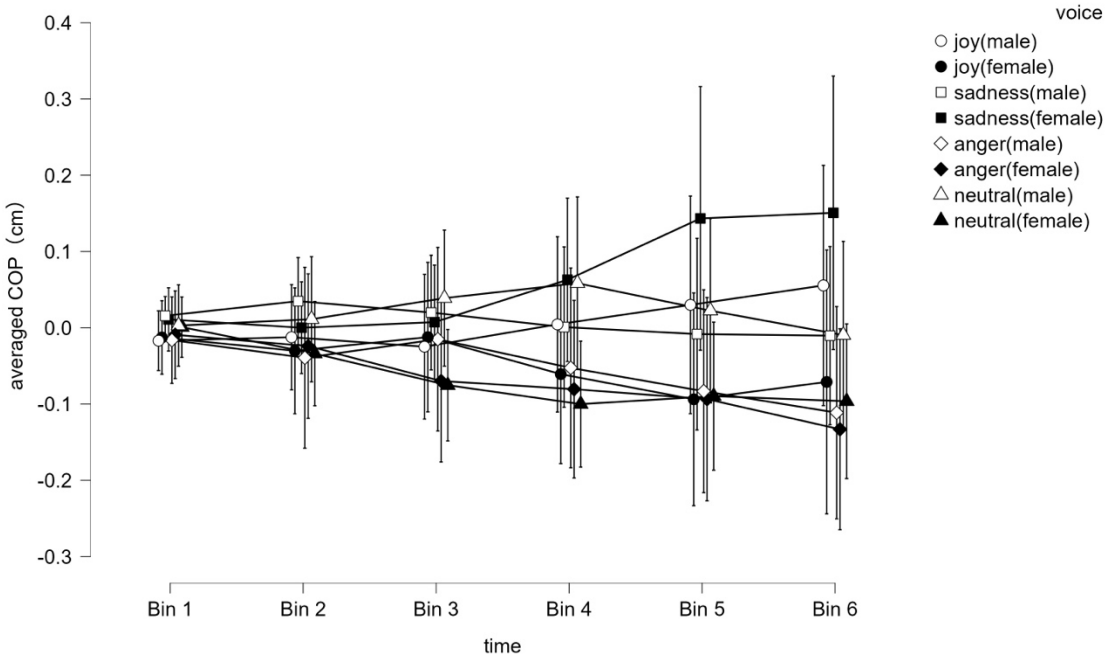
This study was approved by the Institutional Review Board of the Department of Psychology, Hosei University (Approval No. 24-0091).

Results

Figure 1 presents the results of this analysis. The interaction between each vocal stimulus and time was significant (F (35, 805) = 1.717, p = 0.007, η^2 = 0.02). The simple main effect test was significant for sadness (female) (F (5) = 2.609, p = 0.028), anger (female) (F (5) = 2.662, p = 0.026), and neutrality (female) (F (5) = 2.524, p = 0.033); however, the multiple comparison results were not significant. As Figure 1 illustrates, sadness (female) exhibits a movement that gradually approaches, whereas anger (female) and neutrality (female) exhibit a movement that gradually moves in the avoidance direction. Upon visual inspection, anger (males) and joy (females) also appeared to move in the avoidance direction; however, no significant differences were observed for these voices.

Figure 1

Averaged COPs Shift in Eight Types of Vocal Stimuli



Note. The error bars represent 95% confidence intervals.

Discussion and Conclusion

The reason movements approaching sad (female) were observed may be that they induced movements to "help." In contrast, avoidance in response to anger (females) showed results similar to those derived by Kimura et al. (2020), in which avoidance movements were observed when viewing negative videos. Although Kimura et al. (2020) identified no movement in response to neutral videos, this study observed avoidance even with neutral videos (females). This suggests either the presence of negative emotional elements in neutral (female) or the possibility that hearing the voices themselves induced avoidance movements, similar to how Hiraoka et al. (2019) observed avoidance movements with all voices.

Interestingly, interactions were observed for sadness (female), anger (female), and neutrality (female). However, it is noteworthy that the difference was not between positive and negative but rather between avoidance and approach within the same negative category. However, no significant differences were observed for other voices. Specifically, although

changes were observed in the female-negative voices, no significant differences were identified in the male-negative voices.

Two possible explanations can be considered. First, individual differences, such as sensitivity to emotions or voices, may have affected the change in the COP when listening to emotional voices. Second, there may have been a mismatch between the intended emotions in the stimuli and the participants' perceptions of those emotions. Future studies should examine the relationship between individual differences and evaluate how the emotional content of stimuli is perceived.

Acknowledgments

The authors thank the university students who participated in this study.

Declaration of Generative AI and AI-Assisted Technologies in the Writing Process

The author declares that DeepL Translator, an AI-assisted Technologies, was used to translate Japanese into English. The usage was limited to translate accurately and clearly. The author further declares that, apart from DeepL Translator, no other AI or AI-assisted technologies have been used to generate content in writing the manuscript. The ideas, design, procedures, findings, analyses, and discussion are originally written and derived from careful and systematic conduct of the research.

References

Adegawa, A., & Watanabe, Y. (2021). Jidouki ni okeru koe ni yoru kanjyourikai no hattatsu to kanjyou conpitensu no kanren [Development of emotional literacy: Evidence from emotional understanding through voice]. *Proceedings of the 63rd annual meeting of the Japanese Association of Educational Psychology*, 142. https://doi.org/10.20587/pamjaep.63.0_142

Hamana, M., & Haryu, E. (2015). Developmental changes in the semantic domains of emotional words in young children. *The Japanese Journal of Developmental Psychology*, *26*(1), 46–55. https://doi.org/10.11201/jjdp.26.46

Hiraoka, D., Ooishi, Y., Mugitani, R., & Nomura, M. (2019). Differential effects of infant vocalizations on approach-avoidance postural movements in mothers. *Frontiers in Psychology*, *10*, Article 1378. https://doi.org/10.3389/fpsyg.2019.01378

Ikeda, S., & Haryu, E. (2016). Young children's judgments of speaker affect: The developmental mechanism of the appearance and disappearance of lexical bias. *Cognitive Studies: Bulletin of the Japanese Cognitive Science Society*, *23*(1), 49–64. https://doi.org/10.11225/jcss.23.49

Kimura, T., Ueyama, N., & Katayama, J. (2020). Assessment of emotional state by measuring pressure exerted on the seat. *Japanese Society for Physiological Psychology and Psychophysiology, 38*(3), 161–168. https://doi.org/10.5674/jjppp.2001br

Sawada, H., Watanabe, Y., Satou, S., & Oomori, M. (2024). Onsei wo toushita tayouna kanjyourikai no hattatsu [Development of emotional understanding through vocal cues]. *Proceedings of the 66rd annual meeting of the Japanese Association of Educational Psychology*, 522. https://doi.org/10.20587/pamjaep.66.0_522

Watanabe, Y., Takizawa, Y., & Adegawa, A. (2024). The development of emotional prosody recognition among Japanese preschool children. *Bulletin of Faculty of Letters Hosei University*, *88*, 79–98. https://doi.org/10.15002/00030893

Contact email: junko.yakushiji.4a@stu.hosei.ac.jp