



Singing ability is related to vocal emotion recognition: Evidence for shared sensorimotor processing across speech and music

Emma B. Greenspon¹ · Victor Montanaro¹

Accepted: 3 November 2022 / Published online: 15 November 2022
© The Psychonomic Society, Inc. 2022

Abstract

The ability to recognize emotion in speech is a critical skill for social communication. Motivated by previous work that has shown that vocal emotion recognition accuracy varies by musical ability, the current study addressed this relationship using a behavioral measure of musical ability (i.e., singing) that relies on the same effector system used for vocal prosody production. In the current study, participants completed a musical production task that involved singing four-note novel melodies. To measure pitch perception, we used a simple pitch discrimination task in which participants indicated whether a target pitch was higher or lower than a comparison pitch. We also used self-report measures to address language and musical background. We report that singing ability, but not self-reported musical experience nor pitch discrimination ability, was a unique predictor of vocal emotion recognition accuracy. These results support a relationship between processes involved in vocal production and vocal perception, and suggest that sensorimotor processing of the vocal system is recruited for processing vocal prosody.

Keywords Singing accuracy · Music ability · Emotion recognition · Prosody

Introduction

Human language relies on emotional cues that are defined by a number of non-verbal acoustic features, including pitch, timbre, tempo, loudness, and duration (Coutinho & Dibben, 2013). Prosodic features such as fluctuations in vocal pitch and loudness have been linked to physiological responses associated with the emotion that is being expressed in both speech and music (Juslin & Laukka, 2003; Scherer, 2009). According to arousal-based and multi-component theories of emotion, these physiological changes underlie emotion appraisal (James, 1884; Scherer, 2009), and, therefore, physiological arousal may reflect one possible pathway by which vocal cues can convey information to a listener about a speaker's internal state. Furthermore, during in-person interactions, vocal cues are closely coupled with changes in facial behavior (Yehia et al., 1998), reflecting the dynamic and multimodal nature of emotion cues during conversation. Relatedly, automatic mimicry of facial gestures occurs when processing emotional speech and singing (Livingstone et al.,

2009; Stel & van Knippenberg, 2008) and has been linked to emotion recognition (Stel & van Knippenberg, 2008).

Recognition of vocal prosody has also been shown to relate to music background (for a review, see Nussbaum & Schweinberger, 2021). For instance, Dmitrieva et al. (2006) found that musically gifted children showed enhanced vocal emotion recognition compared to age-matched non-musicians. This effect varied by age group, with the largest difference reserved for the youngest group (7–10 years old), which may suggest that early music experience facilitates socio-cognitive development (Gerry et al., 2012). Fuller et al. (2014) reported that effects of musical experience persist in adulthood, with adult musicians exhibiting better vocal emotion recognition than adult non-musicians, and this effect held even under degraded listening conditions. In line with Fuller et al. (2014), Lima and Castro (2011) found that musicians are better at recognizing emotions in speech than non-musicians, even when controlling for other variables like general cognitive abilities and personality traits. To address the directionality of the musician effect, Thompson et al. (2004) and Good et al. (2017) used early music interventions with children. Thompson and colleagues (2004) found that children with musical training in piano, but not voice, recognized vocal emotion more accurately than children without musical training. Similarly, Good et al. (2017) found that children with cochlear implants showed enhanced vocal emotion

✉ Emma B. Greenspon
egreensp@monmouth.edu

¹ Department of Psychology, Monmouth University, West Long Branch, NJ, USA

recognition after musical training in piano compared to a control group that received training in painting.

However, a role of musical experience in vocal prosody processing has not been consistently demonstrated in previous research. For instance, in contrast to Thompson et al. (2004) study, Trimmer and Cuddy (2008), who used the same battery as Thompson et al. (2004), reported that musical training did not account for individual differences in vocal emotion recognition. In that same study, emotional intelligence, on the other hand, was a reliable predictor of vocal emotion recognition, but did not reliably relate to years of musical training (Schellenberg, 2011; cf. Petrides et al., 2006). In addition, Dibben et al. (2018) found an effect of musical training on emotion recognition in music, but not speech.

If musical experience has a role in processing vocal prosody, then one could expect individuals with poor musical abilities to exhibit impairments in recognizing vocal emotion. This claim was addressed by Thompson et al. (2012) and Zhang et al. (2018), who found that individuals with congenital amusia, a deficit in music processing, exhibited lower sensitivity to vocal emotion relative to individuals without amusia. In order to build on work demonstrating that vocal emotion recognition varies by musical ability, the current study was designed to address the role of musical ability in processing vocal emotion using a musical task (i.e., singing) that recruits a shared effector system with speech production.

In order to sing a specific pitch with one's voice, a singer must be able to accurately associate a perceptual representation of the target pitch with the exact motor plan of the vocal system that would produce that pitch. As such, singing is a vocal behavior that reflects sensorimotor processing. Previous work on individual differences in singing ability has found that although inaccurate singing can exist without impaired pitch perception (Pfordresher & Brown, 2007), pitch perception has been shown to correlate with pitch imitation ability (Greenspon & Pfordresher, 2019), with stronger associations observed across singing performance and performance on perceptual measures that assess higher-order musical representations (Pfordresher & Nolan, 2019). Although inaccurate singers can show impairment in matching pitch with their voice, but not when matching pitch using a tuning instrument (Demorest, 2001; Demorest & Clements, 2007; Hutchins & Peretz, 2012; Hutchins et al., 2014), these individuals exhibit similar vocal ranges to accurate singers, non-random imitation performance, and have intelligible speech production, suggesting that these singers express at least some degree of vocal-motor precision (Pfordresher & Brown, 2007). While neither a purely perceptual nor motoric account may be able to fully explain individual differences in singing ability, behavioral studies measuring auditory imagery, a mental process that

recruits both perceptual and motor planning areas of the brain (Herholz et al., 2012; Lima et al., 2016), have supported a sensorimotor account of inaccurate singing (Greenspon et al., 2017; Greenspon et al., 2020; Greenspon & Pfordresher, 2019; Pfordresher & Halpern, 2013).

It is important to note that the ability to accurately vary vocal pitch is not only a critical feature in singing but also an important dimension for communicating spoken prosody, another vocal behavior relying on sensorimotor processing (Aziz-Zadeh et al., 2010; Banissy et al., 2010; Pichon & Kell, 2013). Previous neuroimaging work has established that vocal prosody production recruits overlapping sensorimotor speech pathways used for vocal prosody perception (Aziz-Zadeh et al., 2010). Furthermore, disrupting these sensorimotor pathways through transcranial magnetic stimulation disrupts one's ability to discriminate non-verbal vocal emotions (Banissy et al., 2010). Complementing this finding, Correia et al. (2019) reported that emotion recognition is associated with individual differences in children's sensorimotor processing. Together, these neuroimaging results suggest a link between vocal prosody perception and the vocal system.

Given that both singing and spoken prosody have been linked to individual differences in sensorimotor processing (Aziz-Zadeh et al., 2010; Pfordresher & Brown, 2007; Pfordresher & Mantell, 2014), it is possible that a similar mechanism that accounts for individual differences in vocal imitation of pitch in the context of singing may also account for individual differences in vocal emotion, as suggested by the Multi-Modal Imagery Association (MMIA) model (Pfordresher et al., 2015), a general model of sensorimotor processing based on multi-modal imagery. Such a claim is supported by neuroimaging research that consistently demonstrates that motor planning regions are recruited during auditory imagery for both speech and music (for a review, see Lima et al., 2016). A shared sensorimotor network for singing and vocal emotion also aligns with predictions made by the OPERA hypothesis in which overlapping brain networks for music and speech are proposed to account for the facilitatory effects of music processing on speech processing (Patel, 2011, 2014). Furthermore, behavioral studies support evidence for at least partially shared processes involved in vocal production of speech and song (Christiner & Reiterer, 2013, 2015; Christiner et al., 2022), and have shown that inaccurate imitators of pitch in speech tend to also show impairments in imitating pitch in song (Mantell & Pfordresher, 2013; Wang et al., 2021).

In addition to studies on vocal production, behavioral results have supported the role of vocal pitch perception in speech processing. In a study conducted by Schelinski and von Kriegstein (2019), individuals who were better at discriminating vocal pitch tended to also be better at recognizing

vocal emotion. One disorder that has been linked to deficits in vocal emotion recognition is autism spectrum disorder (ASD; Globerson et al., 2015; Schelinski & von Kriegstein, 2019). Individuals with ASD have been found to exhibit impairments in both vocal pitch perception (Schelinski & von Kriegstein, 2019) and imitation of pitch in speech and song (Jiang et al., 2015; Wang et al., 2021), though ASD can exist with unimpaired non-vocal pitch perception (Schelinski & von Kriegstein, 2019). Together, this pattern of findings suggests that emotion recognition may recruit processes involved in the vocal system and that for those who exhibit impaired emotion recognition, these impairments may extend to behaviors involving vocal production and vocal perception.

We addressed the role of sensorimotor processing in vocal prosody perception for the following reasons. First, physiological changes that occur during felt emotion have been shown to influence vocal expression in both speech and song (Juslin & Laukka, 2003; Scherer, 2009), suggesting that vocal cues can provide information about another's internal state. Second, previous work has found that vocal pitch perception is associated with emotion recognition ability (Schelinski & von Kriegstein, 2019) and that impairments in emotion recognition, vocal production, and vocal perception co-occur (Jiang et al., 2015; Schelinski & von Kriegstein, 2019; Wang et al., 2021), suggesting a possible relationship between emotion processing and the vocal system. Third, neuroimaging work has provided evidence that perceiving vocal prosody recruits overlapping sensorimotor networks involved in vocal production (Aziz-Zadeh et al., 2010; Skipper et al., 2017), and that individual differences in these sensorimotor pathways are related to emotion recognition (Correia et al., 2019). For these reasons, we hypothesized that singing ability would relate to vocal emotion recognition accuracy. Spoken pseudo-sentences were used in the vocal emotion recognition task in order to focus on prosodic features while controlling for semantic information (Pell & Kotz, 2011). We assessed singing ability using a singing protocol that has been found to produce comparable assessments of singing accuracy for in-person and online settings (Honda & Pfordresher, 2022). Pitch discrimination ability was measured in order to address whether vocal emotion recognition ability can be accounted for by lower-level pitch processing, and self-reported musical experience was also assessed.

Method

Participants

Seventy-nine undergraduate students at Monmouth University participated in the study for course credit. Four participants were removed from this sample due to problems related to administering the experiment and four additional

participants were removed due to poor performance levels in at least one task that suggested that participants either did not follow instructions in the task or exhibited a deficit in pitch processing.¹ This resulted in a sample of 71 participants (57 female participants, 14 male participants) who were between 18 and 53 years of age ($M = 20.10$, $SD = 4.48$). Music experience ranged from 0 to 18 years ($M = 3.30$, $SD = 4.86$) and 13 participants reported the voice as their primary instrument. Eight participants reported a language other than English as their first language, and all participants reported learning English by the age of eight years.²

Materials

Singing task

Singing accuracy was measured by participants' performances on the pattern pitch imitation task from the Seattle Singing Accuracy Protocol (SSAP; Demorest et al., 2015) in which participants heard and then imitated four-note novel melodies. Melodies comprised pitches that reflected common comfortable female and male vocal ranges based on unpublished data from the SSAP database. For female participants, melodies were centered around a single pitch (A3) that is typically comfortable for female singers. Melodies were presented one octave lower for male participants, with melodies centered around A2, a pitch that is typically comfortable for male singers.

Pitch discrimination task

Participants also completed a modified non-adaptive version of the pitch discrimination task from the SSAP (Demorest et al., 2015), in which participants heard two pitches and determined whether the second pitch was higher or lower than the initial 500-Hz pitch. There were ten comparison pitches: 300 Hz, 350 Hz, 400 Hz, 450 Hz, 475 Hz, 525 Hz, 550 Hz, 600 Hz, 650 Hz, and 700 Hz. Each comparison pitch was presented five times for a total of 50 trials, and trials were presented in a random order.

¹ Three participants were dropped from this sample due to poor recording quality, one participant was dropped due to experimenter error, one participant was dropped for singing in the wrong octave, two participants were dropped due to extreme contour errors in the singing task ($> 3 SD$ from mean), and based on a priori exclusion criteria one participant was dropped for exhibiting chance-level performance (chance = .5 proportion correct) in the pitch discrimination task.

² Four participants reported Spanish as their first language, one participant reported both English and Spanish as their first language, and three participants reported Chinese, Gujarati, or Urdu as their first language.

Vocal emotion recognition task

Vocal emotion recognition was measured with a selection of 12 English-like pseudo-sentence stimuli (e.g., “The rivix jolled the silling”) from Pell and Kotz (2011). Stimuli were pre-recorded by four speakers (two male and two female speakers). Each speaker conveyed six different emotions (neutrality, happiness, sadness, anger, fear, disgust) for three pseudo-sentences for a total set of 72 stimuli (4 speakers \times 3 sentences \times 6 emotions). As such, there were 12 trials per emotion type. Participants were asked to listen to each sentence and identify the target emotion in a six-option forced-choice task. Stimuli were presented in one of two pseudo-randomized orders, ordered so that no speaker, sentence, or emotion appeared consecutively, and no stimulus was presented in the same position in both orders.

Procedure

Participants completed the experiment in a private Zoom session with the experimenter. Once in the session, participants received a link to the study, which was administered through the online platform FindingFive (FindingFive Team, 2019) in Google Chrome on the participants’ own computers. Audio was presented and recorded by participants’ own headphones/speakers and microphone, and participant recordings were saved to the FindingFive server as a compressed (ogg) file. Participants remained in the Zoom session with their audio connected but their video disabled while completing the experiment through FindingFive. Participants were instructed to sit upright in a chair in order to promote good singing posture before completing a vocal warm-up task. For the vocal warm-up task, participants were instructed to sing a pitch that they found comfortable singing followed by the highest pitch and then the lowest pitch that they could sing. Participants then completed the singing task, which involved imitating a novel pitch sequence of four notes for six trials. These trials were preceded by a practice trial. Following the singing task, participants completed a pitch discrimination task, which asked participants to determine whether a second pitch was higher or lower than the first. Participants then completed the vocal emotion recognition task. On each trial of this task, participants listened to a spoken sentence and identified which one out of six emotions was being conveyed through the sentence’s prosody. Participants were then directed to fill out a musical experience and demographics questionnaire. The experiment took approximately 30 minutes to complete.

Data analysis

In order to analyze performance in the singing task, the compressed (ogg) files were first converted to wav files using the file converter FFmpeg (FFmpeg, 2021). Singing accuracy was

then analyzed by extracting the median f_0 for each sung note using Praat (Boersma & Weenink, 2013). For each note, the difference between the sung f_0 and target f_0 was calculated. A correct imitation was defined as a sung pitch within the range of 50 cents above or below the target pitch. An incorrect imitation was defined as any sung pitch outside of the target range. Correct imitations of a sung pitch were coded as 1 and incorrect imitations were coded as 0. Singing accuracy was averaged within a trial and across the six trials of the singing task.³

Music experience was defined based on self-reported number of years of music experience on the participants’ primary instrument. For the pitch discrimination task, responses that correctly identified that the comparison pitch was higher or lower than the target pitch were coded as 1, while all other responses were coded as 0. Due to high performance in this task, we removed trials with large pitch changes (i.e., greater than a 200-cent difference between the target and comparison pitch) to avoid a ceiling effect and analyzed the remaining 20 trials.

In the vocal emotion recognition task, raw hit rates were calculated by coding a response that correctly identified the intended emotion as 1, while all other responses were coded as 0. We also evaluated accuracy by calculating unbiased hit rates (Wagner, 1993), which aligns with procedures for defining unbiased emotion recognition accuracy in Pell and Kotz (2011). For the unbiased hit rates (H_u), a value of 0 indicated that the emotion label was never accurately matched with the intended emotion, and a value of 1 indicated that the emotion label was always accurately matched with the intended emotion. We did not have hypotheses regarding emotion-specific associations across measures, for this reason, accuracy was then averaged across emotion types in order to provide an overall measure of vocal emotion recognition. This was done for both raw and unbiased hit rates. Bivariate correlations and hierarchical linear regression were conducted to evaluate individual differences in vocal emotion recognition accuracy. All proportion data were arcsine square-root transformed for the regression analyses.

Results

The current study addressed whether individual differences in singing accuracy, pitch discrimination ability, or self-reported musical experience could best account for variability in emotion recognition of spoken pseudosentences. Bivariate correlations across all measures and descriptive statistics for each measure are presented in Table 1. Singing accuracy and pitch discrimination accuracy were calculated as the proportion of

³ A measure of relative pitch accuracy was calculated for the singing task in addition to our measure of absolute pitch accuracy. Relative pitch accuracy was strongly correlated with absolute pitch accuracy ($r = .83, p < .05$) and replicated the relationship between singing accuracy and emotion recognition ($r = .20, p < .05$).

Table 1 Bivariate correlations and descriptive statistics

Measure	1	2	3	4	5
1. Emotion Raw Hit Rates	-	.99**	.25*	.17	.02
2. Emotion Unbiased Hit Rates		-	.26*	.19	.02
3. Singing Accuracy			-	.12	.31**
4. Pitch Discrimination				-	.05
5. Music Experience					-
<i>M</i>	.75	.58	.57	.87	3.30
<i>SD</i>	.08	.11	.30	.10	4.86

Note. * $p < .05$, ** $p < .01$ using a one-tailed test of significance

correct responses in each task, vocal emotion recognition accuracy was measured as raw and unbiased hit rates, and music experience was a self-reported measure of the number of years participants played their primary instrument. Bivariate correlations between predictors and recognition accuracy for different emotion types are presented in the Appendix.

Given the similar pattern observed for both raw and unbiased hit rates shown in Table 1, the remaining analyses focus on unbiased hit rates to measure vocal emotion recognition accuracy while controlling for response bias. As shown in Fig. 1, there was a significant correlation between singing accuracy and unbiased hit rates for vocal emotion recognition such that individuals who were more accurate at imitating pitch tended to be better at recognizing vocal emotion than less accurate singers. In contrast, pitch discrimination ($p = .06$) and self-reported musical experience ($p = .43$) were not correlated with vocal emotion recognition. In addition to an association with vocal emotion recognition, unsurprisingly, singing accuracy was also positively correlated with self-reported musical experience ($p < .01$).

We next conducted a three-step hierarchical linear regression with singing accuracy, pitch discrimination accuracy, and

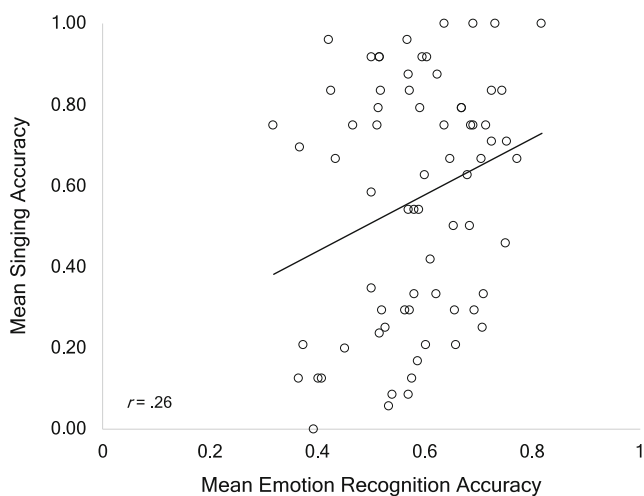


Fig. 1 Bivariate correlation between singing accuracy and vocal emotion recognition

self-reported musical experience as predictor variables and unbiased hit rates for vocal emotion recognition as the dependent variable. Predictors were ordered such that theoretically relevant predictors or predictors that have been previously shown to relate to vocal emotion recognition (Correia et al., 2022; Globerson et al., 2013) were entered before the hypothesized predictor of primary interest (i.e., singing accuracy). As shown in Table 2, only singing accuracy predicted emotion recognition performance above and beyond the other predictors. Alternative orderings of the predictor variables in the model produced the same pattern of results.

Discussion

The current study was designed to address how individual differences in sensorimotor processes pertaining to the vocal system, as measured by singing accuracy, may account for a facilitatory effect of music experience on speech processing. Correlational analyses revealed that singing accuracy was related to vocal emotion recognition and music experience, but neither music experience nor pitch discrimination ability were related to general vocal emotion recognition. Of particular importance to the current study, we observed that singing accuracy was a unique predictor of general vocal emotion recognition ability when controlling for pitch discrimination ability and self-reported musical experience.

We interpret the association between singing accuracy and vocal emotion recognition as evidence for the role of sensorimotor processing in vocal prosody perception. This explanation is motivated by evidence from previous research that inaccurate singing is linked to a sensorimotor deficit (Greenspon et al., 2017; Greenspon et al., 2020; Greenspon & Pfordresher, 2019; Pfordresher & Brown, 2007; Pfordresher & Halpern, 2013; Pfordresher & Mantell, 2014) and that vocal prosody recognition is related to individual differences in sensorimotor processing (Correia et al., 2019). Furthermore, based on our evidence that singing ability, but not self-reported musical experience, is a unique predictor of general vocal emotion recognition, this finding suggests that

Table 2 Three-step hierarchical regression model predicting emotion recognition accuracy

Step/Predictor	Step 1 β	Step 2 β	Step 3 β
Music Experience	.04	.03	-.08
Pitch Discrimination		.18	.15
Singing Accuracy			.31*
Adjusted R^2	-.01	.006	.08*
<i>F</i> for ΔR^2		2.55	6.56*

Note. * $p < .05$, reporting standardized regression estimates

sensorimotor processes involved in spoken prosody may reflect an effector-specific and dimension-specific network of the vocal system recruited for processing pitch in both speech and song. Importantly, a sensorimotor network for processing vocal pitch aligns with the domain general framework of the MMIA model, which is a model accounting for individual differences in sensorimotor processes originally established to account for variability in vocal pitch imitation (Pfordresher et al., 2015). In support of a domain-general effect of sensorimotor processing, previous research has shown that individuals who tend to be poor at imitating pitch in song also tend to be poor at imitating pitch in speech (Liu et al., 2013; Mantell & Pfordresher, 2013; cf. Yang et al., 2014). Furthermore, the sensorimotor account of the relationship between singing accuracy and vocal emotion recognition in the current study is also compatible with the framework proposed by the OPERA hypothesis (Patel, 2011, 2014), in which musical processing is expected to facilitate speech processing for tasks that recruit shared networks involved in both music and speech.

In line with the current results, other studies that have relied on self-report measures of music experience have shown that although emotional intelligence, personality, and age relate to vocal emotion perception, musical training does not (Dibben et al., 2018; Trimmer & Cuddy, 2008). However, studies focused on group comparisons between musicians and non-musicians (Dmitrieva et al., 2006; Fuller et al., 2014; Lima & Castro, 2011; Thompson et al., 2004) and musical training interventions (Good et al., 2017; Thompson et al., 2004) have reported enhanced vocal emotion processing for musically trained individuals. Relatedly, comparisons between individuals with and without a musical impairment (i.e., congenital amusia) reveal that individuals with amusia tend to also exhibit poor vocal emotion perception (Thompson et al., 2012) and that these impairments extend to individuals with tonal language experience (Zhang et al., 2018). Given that amusia has been linked to a deficit specific to pitch processing (Ayotte et al., 2002), one possible explanation for these findings is that individual differences in pitch processing may account for variability in vocal emotion recognition. However, in the current study, pitch discrimination was not a unique predictor of overall vocal emotion recognition. This finding aligns with previous research, which has shown that vocal pitch perception is related to vocal emotion recognition ability; however, pitch perception for non-vocal pitch is not (Schelinski & von Kriegstein, 2019). Complementing these findings, previous research has shown that ASD, which has been linked to difficulty in emotion recognition (Globerson et al., 2015; Schelinski & von Kriegstein, 2019), has also been linked to impairments in vocal perception and vocal production (Jiang

et al., 2015; Schelinski & von Kriegstein, 2019; Wang et al., 2021). Furthermore, neuroimaging research has shown that overlapping neural resources are recruited for both vocal production and perception (Aziz-Zadeh et al., 2010; Skipper et al., 2017), including activity in the inferior frontal gyrus (Aziz-Zadeh et al., 2010; Pichon & Kell, 2013). Interestingly, Aziz-Zadeh et al. (2010) reported that activity in this region during prosody perception correlated with self-reported affective empathy scores (see also Banissy et al., 2012), suggesting a possible link between vocal emotion processing and affective empathy.

In addition to a sensorimotor account of the relationship between singing accuracy and vocal emotion recognition, we also consider whether this relationship can be conceptualized as reflecting individual differences in how auditory information is being prioritized by the listener. In support of this alternative account, Atkinson et al. (2021) have found that listeners can prioritize auditory information when that information is deemed valuable. Furthermore, Sander et al. (2005), who used a dichotic listening task in which participants were instructed to identify a speaker's gender, report that different brain networks are recruited when participants are attending or not attending to angry prosody. Therefore, it may be the case that individuals who are better singers may be better than less accurate singers at prioritizing prosodic cues such as pitch, given that pitch is an important acoustic feature for both spoken prosody and musical performance. This claim aligns with findings from Greenspon and Pfordresher (2019), who found that pitch short-term memory, pitch discrimination, and pitch imagery were unique predictors of singing accuracy, but verbal measures were not. In the current study, participants in the final sample exhibited high levels of pitch discrimination accuracy, suggesting that these individuals did not have difficulty prioritizing pitch information. Furthermore, singing accuracy was a unique predictor of average emotion recognition scores when controlling for individual differences in pitch discrimination ability. However, a limitation of the current study is that pitch perception was measured using a non-adaptive pitch discrimination task with sine wave tones, and therefore cannot address the degree to which individual differences in vocal pitch perception or higher order musical processes involved in melody perception may contribute to the current findings, which are questions that should be addressed in future work.

When considering the results of the current study with respect to task modality, our findings suggest that when assessing musical processes using production and perception-based tasks, the production-based task is a stronger predictor of vocal emotion recognition than the perception-based task. This finding builds on the work by Correia et al.

(2022), who found that perceptual musical abilities (see also Globerson et al., 2013) and verbal short-term memory were both unique predictors of vocal emotion recognition, but musical training was not. However, one limitation of the current study is that only prosody perception, not production, was measured. Therefore, future research is needed to clarify whether individual differences in prosody production relate to singing ability, as found for vocal prosody perception in the current study.

Although the current study focused on general vocal emotion recognition, previous work on vocal expression of emotion suggests that different emotions can be signaled through specific acoustic features, such as variations in pitch contour (Banse & Scherer, 1996; Frick, 1985), and that these cues communicate emotions in both speech and music (Coutinho & Dibben, 2013; Juslin & Laukka, 2003). In addition to being characterized by different acoustic profiles, basic emotions such as anger, disgust, fear, happiness, and sadness have been found to also reflect differences in accuracy and processing time (Pell & Kotz, 2011). For these reasons, we also explored whether singing accuracy, pitch discrimination, and music experience predicted vocal emotion recognition for specific emotions, as discussed in the Appendix. Although all correlations between singing accuracy and vocal emotion recognition showed a positive association, only correlations involving recognition accuracy for sentences portraying fear and sadness reached significance. Correlations between pitch discrimination accuracy and vocal emotion recognition were more variable, with correlations for anger and disgust showing negative, albeit non-significant, relationships. However, pitch discrimination accuracy did positively correlate with vocal emotion recognition for sentences portraying fear, happiness, and neutral emotion. In contrast, we did not find any significant correlations between self-reported musical training and vocal emotion recognition. The emotion-specific pattern reported for these correlations aligns with neuroimaging work that has found emotion-specific neural signatures that are related across different modalities (Aubé et al., 2015; Saarimäki et al., 2016). Furthermore, neuroimaging research has also found that neural responses for specific emotions differ based on musical training with musicians showing different levels of neural activation than non-musicians when listening to spoken sentences portraying sadness (Park et al., 2015). In addition, vocal expression of basic emotions has also been shown to be influenced by physiological changes associated with emotional reactions (Juslin & Laukka, 2003; Scherer, 2009). As such, one pathway by which vocal prosody in speech and song may communicate emotional states of a vocalist is through the association between vocal cues and physiological responses. Such a claim aligns with physiological-based and multi-

component models of emotion processing (James, 1884; Scherer, 2009).

In sum, results of the current study address the degree to which musical ability is associated with processing vocal prosody using a musical production-based singing task that recruits the same effector system as speech. Regression analyses revealed that singing accuracy was the only unique predictor of average spoken prosody recognition, when controlling for pitch discrimination accuracy and self-reported musical experience. Together, our results support sensorimotor processing of the vocal system as a possible mechanism for the facilitatory effects of musical ability on speech processing.

Appendix

We evaluated whether vocal emotion accuracy for different emotion types in the current study replicated the effect of emotion type reported in Pell and Kotz (2011). A one-way repeated-measures ANOVA on unbiased hit rates in the vocal emotion recognition task revealed a main effect of emotion type, $F(5, 350) = 77.16, p < .05$. Descriptive statistics for each emotion (Anger, Disgust, Fear, Happy, Sad, and Neutral) are shown in Appendix Table 3. We conducted pairwise contrasts using a Holm-Bonferroni correction to evaluate differences between emotion types. It is important to note that Pell and Kotz (2011) used a gating procedure whereas the current study used only the full presentation of each sentence (i.e., gate 7), therefore our discussion focuses on the results Pell and Kotz (2011) reported for later gates of the stimuli. We replicated the pattern that fear was recognized with the highest accuracy compared to all other emotions (all $p < .001$) and disgust was recognized with the lowest accuracy compared to all other emotion types (all $p < .001$). In addition, we replicated the finding that accuracy for sentences intended to convey happy emotion were not statistically different from accuracy for sentences intended to convey sad ($p = .17$) nor neutral emotion ($p = .17$).

We next addressed whether singing accuracy, pitch discrimination, and music experience were reliably associated with recognition accuracy for each emotion type. As shown in Appendix Table 3, singing accuracy was positively related to emotion recognition for sentences intended to convey fear and sadness. Correlations between singing accuracy and other emotion types were also positive, but did not reach statistical significance. As found for singing accuracy, pitch discrimination accuracy was positively related to vocal emotion recognition for sentences intended to convey fear. In addition, pitch

Table 3 Descriptive statistics and bivariate correlations between emotion types and predictors

Emotion Type	Unbiased Hit Rates <i>M (SD)</i>	Singing Accuracy <i>r</i>	Pitch Discrimination <i>r</i>	Music Experience <i>r</i>
Anger	.65 (.15)	.07	-.01	-.07
Disgust	.35 (.18)	.10	-.04	.03
Fear	.76 (.15)	.33**	.30**	-.01
Happy	.58 (.19)	.19	.20*	.02
Sad	.54 (.14)	.26*	.14	.14
Neutral	.62 (.15)	.17	.22*	-.01

Note. * $p < .05$, ** $p < .01$ using a one-tailed test of significance. Singing and pitch discrimination accuracy were assessed as the proportion of correct responses in each task, vocal emotion recognition accuracy was assessed as unbiased hit rates, and music experience used a self-report measure expressed in years

discrimination was positively related to emotion recognition scores for sentences intended to convey happiness and neutral emotion. Unlike the associations found with singing accuracy, associations between pitch discrimination and emotion recognition for different emotion types were not consistently in a positive direction. Finally, correlations between self-reported musical experience and emotion recognition also did not show consistently positive associations and did not reach statistical significance for any emotion type.

Acknowledgements The authors would like to thank Marc D. Pell for the stimuli in the vocal emotion recognition task, and Odalys A. Arango, Arelis B. Bernal, Maryam Ettayebi, Joseph LaBarbera, Katherine R. Rivera, Sydney P. Squier, and Adriana A. Zefutue for their assistance with data collection.

Open practices statement We have provided information on participant selection for the final sample, study design, and data analysis. Data for this study is available at (https://osf.io/wa56e/?view_only=0080fadd74274c05b0c5dc13d92b887b). The experiment was not pre-registered.

References

- Atkinson, A. L., Allen, R. J., Baddeley, A. D., Hitch, G. J., & Waterman, A. H. (2021). Can valuable information be prioritized in verbal working memory? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 47(5), 747–764. <https://doi.org/10.1037/xlm0000979>
- Aubé, W., Angulo-Perkins, A., Peretz, I., Concha, L., & Armony, J. L. (2015). Fear across the senses: brain responses to music, vocalizations and facial expressions. *Social Cognitive and Affective Neuroscience*, 10(3), 399–407.
- Ayotte, J., Peretz, I., & Hyde, K. (2002). Congenital amusia: A group study of adults afflicted with a music-specific disorder. *Brain*, 125(2), 238–251. <https://doi.org/10.1093/brain/awf028>
- Aziz-Zadeh, L., Sheng, T., & Gheytanchi, A. (2010). Common premotor regions for the perception and production of prosody and correlations with empathy and prosodic ability. *PLoS One*, 5(1), e8759.
- Banissy, M. J., Sauter, D. A., Ward, J., Warren, J. E., Walsh, V., & Scott, S. K. (2010). Suppressing sensorimotor activity modulates the discrimination of auditory emotions but not speaker identity. *Journal of Neuroscience*, 30(41), 13552–13557.
- Banissy, M. J., Kanai, R., Walsh, V., & Rees, G. (2012). Inter-individual differences in empathy are reflected in human brain structure. *Neuroimage*, 62(3), 2034–2039.
- Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology*, 70(3), 614–636.
- Boersma, P., & Weenink, D. (2013). Praat: doing phonetics by computer (Version 5.4.09). [Software] Available from <http://www.praat.org/>.
- Christiner, M., & Reiterer, S. M. (2013). Song and speech: Examining the link between singing talent and speech imitation ability. *Frontiers in Psychology*, 4, 874. <https://doi.org/10.3389/fpsyg.2013.00874>
- Christiner, M., & Reiterer, S. M. (2015). A Mozart is not a Pavarotti: singers outperform instrumentalists on foreign accent imitation. *Frontiers in Human Neuroscience*, 9, 482. <https://doi.org/10.3389/fnhum.2015.00482>
- Christiner, M., Bernhofs, V., & Groß, C. (2022). Individual Differences in Singing Behavior during Childhood Predicts Language Performance during Adulthood. *Languages*, 7, 72.
- Correia, A. I., Branco, P., Martins, M., Reis, A. M., Martins, N., Castro, S. L., & Lima, C. F. (2019). Resting-state connectivity reveals a role for sensorimotor systems in vocal emotional processing in children. *NeuroImage*, 201, 116052.
- Correia, A. I., Castro, S. L., MacGregor, C., Müllensiefen, D., Schellenberg, E. G., & Lima, C. F. (2022). Enhanced recognition of vocal emotions in individuals with naturally good musical abilities. *Emotion*, 22(5), 894–906.
- Coutinho, E., & Dikken, N. (2013). Psychoacoustic cues to emotion in speech prosody and music. *Cognition & Emotion*, 27(4), 658–684. <https://doi.org/10.1080/02699931.2012.732559>
- Demorest, S. M. (2001). Pitch-matching performance of junior high boys: A comparison of perception and production. *Bulletin of the Council for Research in Music Education*, 63–70.
- Demorest, S. M., & Clements, A. (2007). Factors influencing the pitch-matching of junior high boys. *Journal of Research in Music Education*, 55(3), 190–203.
- Demorest, S. M., Pfordresher, P. Q., Bella, S. D., Hutchins, S., Loui, P., Rutkowski, J., & Welch, G. F. (2015). Methodological perspectives on singing accuracy: An introduction to the special issue on singing accuracy (part 2). *Music Perception: An Interdisciplinary Journal*, 32(3), 266–271. <https://doi.org/10.1525/mp.2015.32.3.266>
- Dikken, N., Coutinho, E., Vilar, J. A., & Estévez-Pérez, G. (2018). Do individual differences influence moment-by-moment reports of

- emotion perceived in music and speech prosody? *Frontiers in Behavioral Neuroscience*, 12, 184. <https://doi.org/10.3389/fnbeh.2018.00184>
- Dmitrieva, E. S., Gel'man, V. Y., Zaitseva, K. A., & Orlov, A. M. (2006). Ontogenetic features of the psychophysiological mechanisms of perception of the emotional component of speech in musically gifted children. *Neuroscience and Behavioral Physiology*, 36(1), 53–62. <https://doi.org/10.1007/s11055-005-0162-6>
- FFmpeg Developers. (2021). ffmpeg tool (Version 4.4). [Software] Available from <http://ffmpeg.org/>
- FindingFive Team. (2019). FindingFive: A web platform for creating, running, and managing your studies in one place. FindingFive Corporation (nonprofit), NJ, USA. <https://www.findingfive.com>
- Frick, R. W. (1985). Communicating emotions: The role of prosodic features. *Psychological Bulletin*, 97(3), 412–429.
- Fuller, C. D., Galvin, J. J., Maat, B., Free, R. H., & Başkent, D. (2014). The musician effect: Does it persist under degraded pitch conditions of cochlear implant simulations? *Frontiers in Neuroscience*, 8, Article 179. <https://doi.org/10.3389/fnins.2014.00179>
- Gerry, D., Unrau, A., & Trainor, L. J. (2012). Active music classes in infancy enhance musical, communicative and social development. *Developmental Science*, 15(3), 398–407.
- Globerson, E., Amir, N., Golan, O., Kishon-Rabin, L., & Lavidor, M. (2013). Psychoacoustic abilities as predictors of emotion recognition. *Attention, Perception, & Psychophysics*, 75(8), 1,799–1,810. <https://doi.org/10.3758/s13414-013-0518-x>
- Globerson, E., Amir, N., Kishon-Rabin, L., & Golan, O. (2015). Prosody recognition in adults with high-functioning autism spectrum disorders: From psychoacoustics to cognition. *Autism Research*, 8(2), 153–163.
- Good, A., Gordon, K. A., Papsin, B. C., Nespoli, G., Hopyan, T., Peretz, I., & Russo, F. A. (2017). Benefits of music training for perception of emotional speech prosody in deaf children with cochlear implants. *Ear and Hearing*, 38(4), 455.
- Greenspon, E. B., & Pfordresher, P. Q. (2019). Pitch-specific contributions of auditory imagery and auditory memory in vocal pitch imitation. *Attention, Perception, & Psychophysics*, 81(7), 2473–2481.
- Greenspon, E. B., Pfordresher, P. Q., & Halpern, A. R. (2017). Pitch imitation ability in mental transformations of melodies. *Music Perception: An Interdisciplinary Journal*, 34(5), 585–604.
- Greenspon, E. B., Pfordresher, P. Q., & Halpern, A. R. (2020). The role of long-term memory in mental transformations of pitch. *Auditory Perception & Cognition*, 3(1-2), 76–93.
- Herholz, S. C., Halpern, A. R., & Zatorre, R. J. (2012). Neuronal correlates of perception, imagery, and memory for familiar tunes. *Journal of Cognitive Neuroscience*, 24, 1382–1397. https://doi.org/10.1162/jocn_a_00216
- Honda, C., & Pfordresher, P. Q. (2022). *Remotely collected data can be as good as laboratory collected data: A comparison between online and in-person data collection in vocal production* [Manuscript in revision for publication].
- Hutchins, S. M., & Peretz, I. (2012). A frog in your throat or in your ear? Searching for the causes of poor singing. *Journal of Experimental Psychology: General*, 141(1), 76–97.
- Hutchins, S., Larrouy-Maestri, P., & Peretz, I. (2014). Singing ability is rooted in vocal-motor control of pitch. *Attention, Perception, & Psychophysics*, 76(8), 2522–2530.
- James, W. (1884). What is an emotion? *Mind*, 9(34), 188–205.
- Jiang, J., Liu, F., Wan, X., & Jiang, C. (2015). Perception of melodic contour and intonation in autism spectrum disorder: Evidence from Mandarin speakers. *Journal of Autism and Developmental Disorders*, 45(7), 2067–2075.
- Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin*, 129(5), 770–814.
- Lima, C. F., & Castro, S. L. (2011). Speaking to the trained ear: Musical expertise enhances the recognition of emotions in speech prosody. *Emotion*, 11(5), 1021–1031.
- Lima, C. F., Krishnan, S., & Scott, S. K. (2016). Roles of supplementary motor areas in auditory processing and auditory imagery. *Trends in Neurosciences*, 39(8), 527–542.
- Liu, F., Jiang, C., Pfordresher, P. Q., Mantell, J. T., Xu, Y., Yang, Y., & Stewart, L. (2013). Individuals with congenital amusia imitate pitches more accurately in singing than in speaking: Implications for music and language processing. *Attention, Perception, & Psychophysics*, 75(8), 1783–1798.
- Livingstone, S., Thompson, W. F., & Russo, F. A. (2009). Facial expressions and emotional singing: A study of perception and production with motion capture and electromyography. *Music Perception*, 26, 475–488.
- Mantell, J. T., & Pfordresher, P. Q. (2013). Vocal imitation of song and speech. *Cognition*, 127(2), 177–202. <https://doi.org/10.1016/j.cognition.2012.12.008>
- Nussbaum, C., & Schweinberger, S. R. (2021). Links between musicality and vocal emotion perception. *Emotion Review*, 13(3), 211–224.
- Park, M., Gutyrchik, E., Welker, L., Carl, P., Pöppel, E., Zaytseva, Y., et al. (2015). Sadness is unique: neural processing of emotions in speech prosody in musicians and non-musicians. *Frontiers in Human Neuroscience*, 8, 1049. <https://doi.org/10.3389/fnhum.2014.01049>
- Patel, A. D. (2011). Why would musical training benefit the neural encoding of speech? The OPERA hypothesis. *Frontiers in Psychology*, 2, 1–14.
- Patel, A. D. (2014). Can nonlinguistic musical training change the way the brain processes speech? The expanded OPERA hypothesis. *Hearing Research*, 308, 98–108.
- Pell, M. D., & Kotz, S. A. (2011). On the time course of vocal emotion recognition. *PLoS One*, 6(11), e27256. <https://doi.org/10.1371/journal.pone.0027256>
- Petrides, K. V., Niven, L., & Mouskounti, T. (2006). The trait emotional intelligence of ballet dancers and musicians. *Psicothema*, 18, 101–107.
- Pfordresher, P. Q., & Brown, S. (2007). Poor-pitch singing in the absence of "tone deafness". *Music Perception*, 25(2), 95–115.
- Pfordresher, P. Q., & Halpern, A. R. (2013). Auditory imagery and the poor-pitch singer. *Psychonomic Bulletin & Review*, 20(4), 747–753.
- Pfordresher, P. Q., & Mantell, J. T. (2014). Singing with yourself: Evidence for an inverse modeling account of poor-pitch singing. *Cognitive Psychology*, 70, 31–57.
- Pfordresher, P. Q., & Nolan, N. P. (2019). Testing convergence between singing and music perception accuracy using two standardized measures. *Auditory Perception & Cognition*, 2(1-2), 67–81.
- Pfordresher, P. Q., Halpern, A. R., & Greenspon, E. B. (2015). A mechanism for sensorimotor translation in singing: The Multi-Modal Imagery Association (MMIA) model. *Music Perception: An Interdisciplinary Journal*, 32(3), 242–253.
- Pichon, S., & Kell, C. A. (2013). Affective and sensorimotor components of emotional prosody generation. *Journal of Neuroscience*, 33(4), 1640–1650.
- Saarimäki, H., Gotsopoulos, A., Jääskeläinen, I. P., Lampinen, J., Vuilleumier, P., Hari, R., ... & Nummenmaa, L. (2016). Discrete neural signatures of basic emotions. *Cerebral Cortex*, 26(6), 2563–2573.
- Sander, D., Grandjean, D., Pourtois, G., Schwartz, S., Seghier, M. L., Scherer, K. R., & Vuilleumier, P. (2005). Emotion and attention interactions in social cognition: brain regions involved in processing anger prosody. *Neuroimage*, 28(4), 848–858.
- Schelinski, S., & von Kriegstein, K. (2019). The relation between vocal pitch and vocal emotion recognition abilities in people with autism spectrum disorder and typical development. *Journal of Autism and Developmental Disorders*, 49(1), 68–82.

- Schellenberg, E. G. (2011). Music lessons, emotional intelligence, and IQ. *Music Perception, 29*(2), 185–194. <https://doi.org/10.1525/mp.2011.29.2.185>
- Scherer, K. R. (2009). The dynamic architecture of emotion: Evidence for the component process model. *Cognition and Emotion, 23*(7), 1307–1351.
- Skipper, J. I., Devlin, J. T., & Lametti, D. R. (2017). The hearing ear is always found close to the speaking tongue: Review of the role of the motor system in speech perception. *Brain and Language, 164*, 77–105.
- Stel, M., & van Knippenberg, A. (2008). The role of facial mimicry in the recognition of affect. *Psychological Science, 19*(10), 984–985.
- Thompson, W. F., Schellenberg, E. G., & Husain, G. (2004). Decoding speech prosody: Do music lessons help? *Emotion, 4*(1), 46–64.
- Thompson, W. F., Marin, M. M., & Stewart, L. (2012). Reduced sensitivity to emotional prosody in congenital amusia rekindles the musical protolanguage hypothesis. *Proceedings of the National Academy of Sciences of the United States of America, 109*(46), 19, 027–19,032. <https://doi.org/10.1073/pnas.1210344109>
- Trimmer, C. G., & Cuddy, L. L. (2008). Emotional intelligence, not music training, predicts recognition of emotional speech prosody. *Emotion, 8*(6), 838–849. <https://doi.org/10.1037/a0014080>
- Wagner, H. L. (1993). On measuring performance in category judgment studies of nonverbal behavior. *Journal of Nonverbal Behavior, 17*(1), 3–28.
- Wang, L., Pfordresher, P. Q., Jiang, C., & Liu, F. (2021). Individuals with autism spectrum disorder are impaired in absolute but not relative pitch and duration matching in speech and song imitation. *Autism Research, 14*(11), 2355–2372.
- Yang, W. X., Feng, J., Huang, W. T., Zhang, C. X., & Nan, Y. (2014). Perceptual pitch deficits coexist with pitch production difficulties in music but not Mandarin speech. *Frontiers in Psychology, 4*, 1024. <https://doi.org/10.3389/fpsyg.2013.01024>
- Yehia, H., Rubin, P., & Vatikiotis-Bateson, E. (1998). Quantitative association of vocal-tract and facial behavior. *Speech Communication, 26*(1-2), 23–43.
- Zhang, Y., Geng, T., & Zhang, J. (2018, September 2-6). Emotional prosody perception in Mandarin-speaking congenital amusics. In: *Proceedings of the Annual Conference of the International Speech Communication Association (Interspeech 2018)*, 2196–2200.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.