ASHA American Speech-Language-Hearing Association

JSLHR

Research Article

Relationship Between Dynamic Pitch Sensitivity and Vocal Emotion Recognition in Different Listening Conditions in Older Listeners

Benjamin T. Amartey,[a] Monita Chatterjee,[a] and Pamela Souza[a]

[a] Department of Communication Sciences and Disorders, Northwestern University, Evanston, IL

ARTICLE INFO

Article History:
Received November 13, 2025
Revision received February 9, 2026
Accepted April 11, 2026

Editor-in-Chief: Rachael Frush Holt
Editor: Gavin M. Bidelman

https://doi.org/10.1044/2026_JSLHR-25-00918

ABSTRACT

Purpose: This study examined the relationship between sensitivity to dynamic pitch (i.e., perception of intonation) and vocal emotion recognition in older adults under different listening conditions. We hypothesized that better dynamic pitch sensitivity would be positively associated with accurate recognition of vocal emotions, regardless of age or listening context.
Method: We conducted two related studies. In Study 1, 32 listeners (ages 55–88 years) completed a pitch direction identification task assessing dynamic pitch sensitivity and a vocal emotion recognition task in quiet. In Study 2, 21 listeners (ages 56–88 years) from Study 1 completed the emotion recognition task in babble maskers presented from multiple locations around the listener at two signal-to-noise ratios.
Results: In Study 1, results showed a significant positive relationship between dynamic pitch sensitivity and vocal emotion recognition. Study 2 found a similar relationship between dynamic pitch sensitivity and vocal emotion recognition as was observed in Study 1; however, this relationship did not reach statistical significance. Participants' age was found to negatively predict vocal emotion accuracy across both listening conditions.
Conclusions: Our findings highlight the unique contribution of dynamic pitch sensitivity to individual differences in vocal emotion recognition in older listeners. These findings replicate and extend the literature on emotional communication in aging and suggest that dynamic pitch sensitivity may be one of the mechanisms supporting accurate emotion recognition in aging.

Affective communication is an important aspect of human interactions. In social communication, both what is being communicated and the intentions and emotions underlying it are equally important for successful communication. The affective aspect of communication is beneficial for appropriate social interactions, emotional well-being, and improved quality of life (Adolphs, 2002; Pawełczyk et al., 2021). When an individual has an impaired ability to recognize vocal emotions, it may lead to social withdrawal and confusion (Pawełczyk et al., 2021). Emotional information in human speech is expressed via emotional semantics (words) and/or emotional prosody (the tone of voice). Both channels typically complement each other but may sometimes contradict each other. In such situations, the listener has to decide which channel best captures the emotion the speaker is trying to convey (Kotz & Paulmann, 2011; Schirmer & Kotz, 2006). Therefore, when verbal cues are ambiguous or uninformative, accurate interpretation of prosodic cues becomes important for emotional communication.

Vocal emotions rely on a variety of acoustic cues such as voice pitch, vocal timbre, speaking rate, and loudness (Banse & Scherer, 1996; Scherer, 2003). While all of these cues are important, the most prominent cue for conveying emotions in speech appears to be pitch contour, defined as dynamic changes in the fundamental frequency

Correspondence to Benjamin T. Amartey: benjamin.amartey@northwestern.edu. ***Disclosure:*** *The authors have declared that no competing financial or nonfinancial interests existed at the time of publication.*

(*F*0; Juslin & Laukka, 2003; Scherer, 2003). Dynamic pitch variations may serve as unique acoustic characteristics for vocal emotion, where an increase in pitch and loudness is typically associated with "happiness" and a decrease in pitch is associated with "sadness" (Juslin & Laukka, 2003; Scherer, 2003). Therefore, recognizing vocal emotions accurately may depend on the ability to track and interpret the dynamic changes in pitch (Globerson et al., 2013; Juslin & Laukka, 2003; Nussbaum et al., 2023). However, when dynamic pitch contours are smoothed or flattened, vocal emotion recognition may be impaired (Lieberman & Michaels, 1962). Given that successful vocal emotion recognition is supported by sensitivity to dynamic pitch information, individual differences in this ability may have consequences for emotional communication.

One determinant of individual ability may be the listener's age. Meta-analytic reviews and individual studies have consistently documented general age-related declines in the ability to identify vocal emotions from prosodic cues (Baglione et al., 2025; Cannon & Chatterjee, 2022; Christensen et al., 2019; Fan et al., 2024). Surprisingly, the declines were not uniform across emotional categories. Recognition of anger and sadness showed the largest age-related deficits, while recognition of happiness was relatively preserved (Ruffman et al., 2008). These deficits are not fully explained by reduced audibility alone (Irino et al., 2024; Rachman & Başkent, 2025) and are often present even among older listeners with normal audiometric thresholds (Cannon & Chatterjee, 2022; Christensen et al., 2019).

Age-related declines in emotion recognition may also affect the amount of importance listeners assign to different acoustic cues. Younger listeners prioritize prosodic cues, particularly pitch information, when making emotional judgments (Dupuis & Pichora-Fuller, 2010). However, older listeners rely more on semantic content, especially when prosodic and semantic channels conflict (Ben-David et al., 2019; Dor et al., 2022, 2024; Lin et al., 2024). It is possible that the shift in cue weighting by older listeners is a compensatory strategy related to the loss of access to prosodic information.

Several explanations have been proposed for age-related deficits in vocal emotion recognition. From a peripheral and sensory perspective, age-related changes in how the auditory system encodes *F*0 may degrade the quality of prosodic information available for higher processing. Declines in sensitivity to temporal fine structure and dynamic pitch cues may reduce access to acoustic features important for conveying emotional information in speech (Anderson et al., 2011; Tremblay et al., 2021). Importantly, despite established group-level patterns (i.e., better performance by younger vs. older adults), there is substantial individual variability in older adults' ability to recognize vocal emotions. Some older individuals perform comparably to younger adults, while others show deficits (Dupuis & Pichora-Fuller, 2015; Paulmann et al., 2008). Such variability motivates examination of the factors, particularly auditory perceptual abilities, that might explain why some older adults are more vulnerable to emotion recognition difficulties than others.

One potential explanation for individual differences in vocal emotion recognition is that some listeners have poor sensitivity to pitch. Older adults demonstrate poor pitch sensitivity as reflected by larger (poorer) *F*0 difference limens (*F*0DLs; Shen et al., 2016; Souza et al., 2011) and declines in subcortical representations of pitch (Clinard et al., 2010). Previous studies suggest that the ability to discriminate *F*0, indexed by *F*0DLs, may support dynamic pitch sensitivity (Chatterjee & Peng, 2008; Deroche et al., 2016; Souza et al., 2011). Here, we examine the relationship between dynamic pitch sensitivity and vocal emotion recognition to determine whether dynamic pitch sensitivity contributes to individual differences in vocal emotion recognition. Establishing this relationship is important for understanding the mechanisms underlying vocal emotion recognition in aging and for identifying potential targets for intervention.

While examining the relationship between dynamic pitch sensitivity and vocal emotion in quiet is an important first step, vocal emotion recognition often occurs in more challenging listening environments. As such, if the relationship between dynamic pitch sensitivity and emotion recognition is to have real-world relevance, it is also important to examine it under conditions characteristic of everyday listening environments. Everyday listening environments, including restaurants, family gatherings, and workplaces, often contain competing sound sources that can interfere with the perception of a target speaker's voice (Mattys et al., 2012; Smeds et al., 2015; Wu et al., 2018). Competing sound sources introduce spectral and temporal interference that may mask pitch variations and reduce the clarity of perceived pitch contours (Song et al., 2011). Several studies have shown that recognition of the emotional content of speech is diminished by noise (Dor et al., 2022; Morgan, 2021; Zhang & Ding, 2022). However, other studies have indicated that emotional cues associated with the variation in pitch are more robust to the effects of noise than other emotional cues (Guo et al., 2022; Kong & Zeng, 2006). The inconsistencies across these studies support the notion that effective extraction of pitch information from noise may be the reason why some listeners are still able to recognize emotions in noisy environments, while others have difficulty. Therefore, we investigated how well the ability to track and interpret dynamic pitch relates to the ability to recognize vocal emotion in noise typical of many daily listening situations.

Another focus of our study is the inability of objective vocal emotion measures to capture all the difficulties

that are associated with emotional communication in everyday listening environments. To address this limitation, we included the Emotional Communication in Hearing Questionnaire (EMO-CHeQ) as a self-report survey, in addition to the behavioral emotion recognition task. Our goal is to assess whether the reported difficulties in emotional communication by listeners are consistent with their performance on the objective measure. The agreement or disagreement between these two different measures may offer insight into the functional impact of emotion recognition in everyday social interactions.

This work addresses two interrelated questions. First, we examine whether dynamic pitch sensitivity predicts individual differences in vocal emotion recognition accuracy. We hypothesize that strong sensitivity to dynamic pitch will be associated with accurate vocal emotion recognition. Second, we investigate whether sensitivity to dynamic pitch supports emotion recognition when listening condition is degraded by surrounding noise. Including both quiet and noise conditions enables examination of how the relationship between dynamic pitch sensitivity and vocal emotion fares across different listening conditions. We also examine whether objective measures of vocal emotion recognition correspond with self-reported measures of emotion perception, as assessed by the EMO-CHeQ.

Study 1

In Study 1, we investigated whether sensitivity to dynamic pitch predicts individual differences in vocal emotion recognition in a quiet environment.

Method

Participants

Thirty-two listeners (18 female, 14 male) aged 55–88 years (M = 70.32, SD = 8.37) participated in Study 1. Participants were eligible if their better-ear four-frequency pure-tone average (PTA; 500, 1000, 2000, and 4000 Hz) fell between 0 and 55 dB HL (see Figure 1). Hearing thresholds were measured for each ear at octave and mid-octave frequencies ranging from 250 to 8000 Hz using standard audiometric procedures in a double-walled sound-treated booth. Mean right-ear PTA was 27.91 dB HL, and mean left-ear PTA was 28.29 dB HL. All participants reported English as their primary language, had no proficiency in tonal languages, and reported fewer than 2 years of formal music training. Cognitive status was screened using the Montreal Cognitive Assessment (Nasreddine et al., 2005), with a cutoff score of 23 or higher, which has been shown to provide good sensitivity and specificity for detecting cognitive impairment in older adults (Carson et al., 2018). Recruitment was conducted through the Hearing Aid Laboratory's volunteer pool and community posts in the Evanston and greater Chicago area. Each participant completed a single session of approximately 2 hr and was compensated for their time. The Northwestern University Institutional Review Board (IRB No. STU00218810) approved the study, and informed consent was obtained from all participants. The participant demographics are presented in Table 1.

Dynamic Pitch Perception

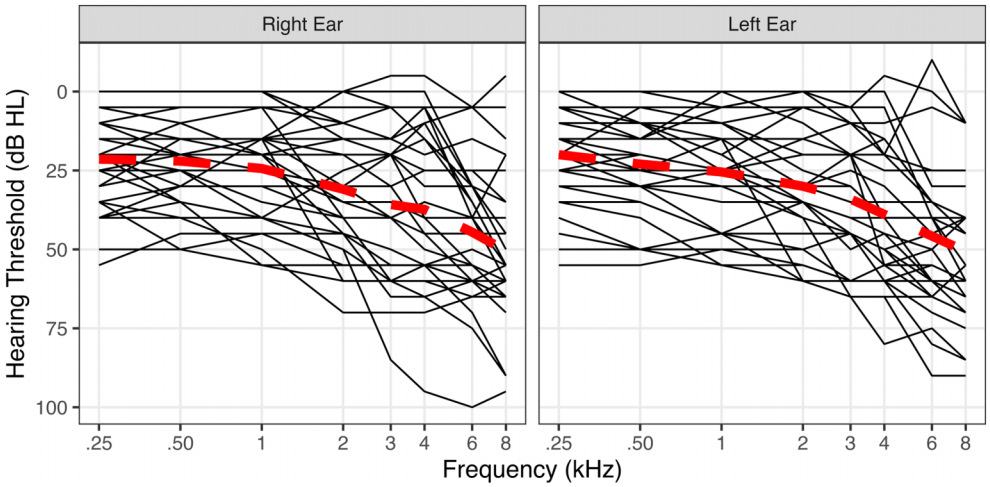
Stimuli

Sensitivity to dynamic pitch (intonation) was measured using a pitch direction identification task. The stimulus set included four diphthongs: /aʊ/, /eɪ/, /ɪ/, and /ɔ/ (as in cow, day, eye, and boy, respectively), created by Green et al. (2002) using a Klatt synthesizer in cascade mode (20 kHz sample rate, parameters specified every 5 msecs). Each diphthong was 620 ms in duration. Formant trajectories were modeled on recordings of a male speaker. For each diphthong, $F0$ varied linearly over the stimulus duration to create rising or falling pitch contours. The $F0$ at the temporal midpoint was 113 Hz (corresponding to a male voice). For the 113 Hz midpoint, the ratio of start-to-end $F0$ varied in 12 logarithmic steps from 1:0.5 (rising: e.g., 80 Hz → 160 Hz) to 1:2.0 (falling: e.g., 160 Hz → 80 Hz). The ratio of start-to-end $F0$ varied in 12 logarithmic steps from 1:0.5 (rising: e.g., 160 Hz → 320 Hz) to 1:2.0 (falling: e.g., 320 Hz → 160 Hz). The choice of diphthongs allowed us to examine listeners' ability to track dynamic pitch changes arising from $F0$ variation in a naturalistic speech token, that is, in the context of formant structure.

Task and Procedure

A MATLAB program controlled the pitch direction identification task. All the stimuli were delivered through an M-Audio interface for digital-to-analog conversion. Testing was conducted in a sound-treated booth using free-field presentation from a Yamaha HS50M loudspeaker positioned at ear level, 1 m, and at 0 ° azimuth from the listener. Presentation level was customized for each listener as follows. First, individual pure-tone thresholds were converted from dB HL to dB SPL (American National Standards Institute, 2018). Next, individual sound-field thresholds were estimated following conversion values from Yost and Killion (1997). If, for a given participant, any one-third-octave band level of the dynamic pitch glides was below 20 dB above the listener's threshold, the stimulus presentation level for that listener was increased until a minimum 20 dB SL threshold was achieved in each one-third-octave band. Final presentation levels ranged from 69 to 75 dB SPL. In each trial, the listeners used a computer mouse to determine whether the contour was a "rise" or "fall." Two base $F0$s were tested (113 and 226 Hz). Each

Figure 1. Right and left ear audiograms from all participants (black lines) in Study 1. The red dashed lines indicate the overall mean threshold.



block represented a talker (male or female), and each participant completed 48 randomized trials (4 diphthongs × 12 *F*0 steps) within each block. Before each block, a practice block of 24 trials was presented, with visual feedback. Each *F*0 (113 and 226 Hz) was then tested in two blocks of 48 trials with no feedback. The talker order was randomized across listeners. Testing was performed unaided. Listeners' responses ("rising" or "falling") were recorded for each trial and later analyzed to estimate dynamic pitch sensitivity.

Vocal Emotion Recognition

Stimuli

Stimuli comprised adult-directed vocal emotion recordings similar to those used in previous studies (Cannon & Chatterjee, 2022; Christensen et al., 2019). The recordings were made from 12 semantically neutral sentences from the Hearing in Noise Test corpus Nilsson et al. (1994), recorded in one of five different emotions by a male and a female talker. There were five emotions (happy, angry, scared, sad, and neutral), making a total of 60 sentences per talker.

Task and Procedure

The experiment was controlled by a custom MATLAB program (Emognition) used in previous studies (Cannon & Chatterjee, 2019, 2022). Participants responded to the task by clicking a button on a computer screen. For each participant, the initial talker (i.e., male or female) was randomly selected, and the talker block was subsequently counterbalanced. Prior to each block, passive training was administered. The training included 10 sentences (two sentences spoken in five emotions) per talker. The sentences used in the passive training were excluded from the main task. During the training session, participants were presented with 10 utterances. Each utterance played was followed by an interval in which the correct emotion associated with the utterance was highlighted on the computer screen among five different emotion options. The training was repeated twice, after which the actual test was administered. The actual test was administered in a five-alternative forced-choice paradigm within a single interval. After each utterance was played, participants were required to choose which of the five emotions was being conveyed in the utterance. The options were stacked vertically as stick-figure faces on the left side of the computer screen. No feedback was given to participants, and repetition of stimuli was prohibited. The actual test consisted of two talker blocks of 60 trials each (5 emotions × 12 sentences). Presentations were conducted in a random order. Participants were allowed to take a brief break between blocks. Participants' responses were recorded as overall percent correct scores and confusion matrices. Presentation levels specific to each listener were used across both the dynamic pitch identification and vocal emotion recognition tasks.

Table 1. Participant demographics for Study 1.

Characteristic	*M*	*SD*
Age (years)	70.32	8.37
MoCA score	27.24	1.63
Right PTA (dB HL)	27.91	15.40
Left PTA (dB HL)	28.29	14.85

Note. MoCA = Montreal Cognitive Assessment; PTA = pure-tone average.

Statistical Analysis

Dynamic Pitch Sensitivity

The dynamic pitch data were visualized by transforming the stimulus *F*0 ratio to a base-10 logarithmic scale to provide a linear distribution of values. A logistic regression

was fitted to each listener's data, where the log-transformed *F*0 ratio served as the predictor variable and the proportion of "falling" responses at each *F*0 ratio served as the outcome. The slope of the fitted function was used as a measure of dynamic pitch sensitivity, where steeper slopes indicate greater sensitivity to pitch direction and shallower slopes indicate less sensitivity.

Vocal Emotion Recognition

Sensitivity (d'), a signal detection measure that distinguishes between perceptual sensitivity and response bias (Macmillan & Kaplan, 1985), was used to quantify emotion recognition performance. Sensitivity (d') for each of the five emotion categories was derived from the confusion matrices. The hit rate for each emotion was determined by the proportion of trials on which that emotion was incorrectly selected when a different emotion was presented. The false alarm, on the other hand, was determined by the proportion of trials during which participants incorrectly selected that emotion when another emotion was presented. The formula used in calculating the d' for each emotion category was

$$d' = \text{z(Hit rate)}–\text{z(False alarm rate)} \quad (1)$$

where z is the inverse of the standard normal cumulative distribution function. Prior to performing z transformation, hit rates of 1.00 were adjusted to 0.9999, and false alarm rates of 0.00 were adjusted to 0.0001 to avoid infinite values (Hautus et al., 2021; Macmillan & Kaplan, 1985). Finally, the d' values for the five emotions were averaged together to provide a single composite measure of emotion recognition sensitivity for each speaker and participant. This averaging approach treats each emotion category as contributing equally to overall recognition ability.

Linear mixed-effects models were constructed using the lme4 package (Bates et al., 2015), with degrees of freedom and p values computed via Satterthwaite's approximation as implemented in lmerTest (Kuznetsova et al., 2017). Models were estimated using maximum likelihood to allow for likelihood ratio tests comparing nested models. Model building proceeded as follows. A baseline model including only a random intercept for participant was first specified to account for repeated measures across speaker conditions. Fixed effects were then added sequentially: dynamic pitch slope (sensitivity to dynamic pitch cues), age (centered), and PTA (centered). PTA did not improve the model fit nor reach significance, so it was removed. Adding the Age × Dynamic Pitch interaction did not improve the fit and was also excluded from the final model. Model comparisons were conducted using chi-square tests of the −2 log-likelihood ratio via the anova() function in R, with lower Akaike information criterion (AIC) and Bayesian information criterion (BIC) values indicating better fit (Burnham & Anderson, 2004). Random slopes for fixed effects were considered but resulted in convergence failures, likely due to insufficient variance in the random effects structure given the sample size. Therefore, the final models retained only random intercepts for participants. Predictors that did not significantly improve model fit ($p > .05$) were removed from the final model. The full model for emotion recognition sensitivity was specified as

$$d' \sim \text{Dynamic pitch slope} + \text{Age} + (1 \mid \text{Participant}) \quad (2)$$

Model assumptions were evaluated through visual inspection of diagnostic plots. Residual histograms and Q–Q plots were examined to assess normality of residuals, and residual-versus-fitted plots were inspected for evidence of heteroscedasticity. Random effects were examined to ensure approximate normality of participant-level intercepts. The models were estimated using the maximum likelihood. Residual histograms, Q–Q plots, and residual-versus-fitted scatter plots confirmed acceptable normality and homoscedasticity. All analyses were conducted in R (Version 4.4.1) using the lme4 package (Bates et al., 2015), with degrees of freedom and p values estimated via Satterthwaite's approximation as implemented in lmerTest (Kuznetsova et al., 2017). Ninety-five percent confidence intervals (CIs) for fixed effects were estimated using parametric bootstrapping with 5,000 resamples (percentile method). Figures were generated using ggplot2 (Wickham, 2016).

Results

Descriptive Statistics

We used data from 32 listeners for the analyses. Table 1 provides descriptive statistics of the primary.

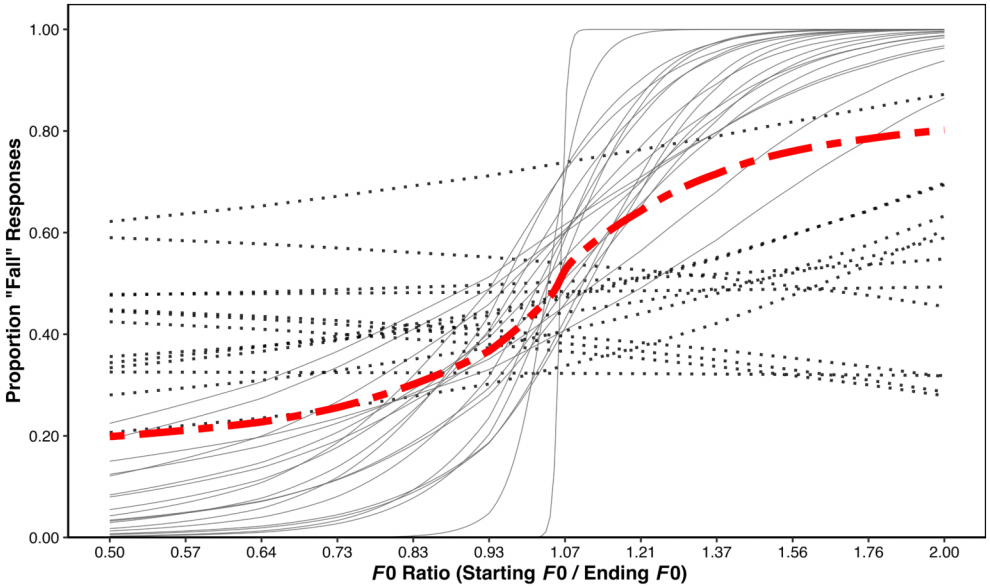
Dynamic Pitch Sensitivity

Psychometric functions based on the individual listener's responses to the dynamic pitch stimuli are shown in Figure 2. In this figure, steeper slopes (solid lines) represent stronger sensitivity to the direction of the pitch changes, while shallow slopes (dotted lines) indicate less sensitivity to the direction of the pitch changes. Listeners' performance showed variability in slope: Some listeners demonstrated steeper slopes, and others had shallower slopes. The group mean was represented by the thick two-dash red line.

Vocal Emotion Recognition

Confusion matrices illustrating general response patterns are shown in Figure 3 for all listeners. Correct identifications are located along the diagonal of each panel and

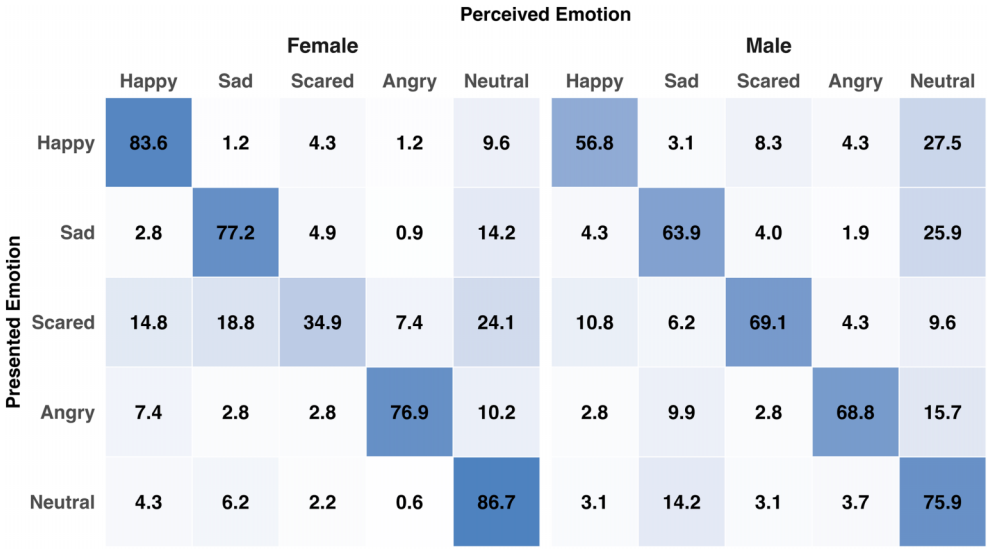
Figure 2. Psychometric functions for dynamic pitch sensitivity in older listeners. The abscissa shows the fundamental frequency (*F*0) ratio between the beginning and end of the diphthong. The ordinate shows the proportion of trials identified as "falling." Individual functions are represented by the solid (steeper slopes) and dotted (shallower) lines. The thick two-dash red line shows the group average. An ideal response pattern would approach 0% "falling" at 0.5, about 50% at 1.0, and 100% at 2.0. Steeper slopes indicate stronger sensitivity to pitch direction, while shallower slopes indicate weaker perception. The figure illustrates clear individual variability: Some listeners produce functions close to the ideal pattern, while others show flatter functions with little sensitivity to dynamic pitch.



show that listeners demonstrated an overall high accuracy in identifying emotions. Accuracy of identification of emotions was lowest for the male talker, and happy utterances were most frequently misidentified as neutral. For the female talker, happy, sad, angry, and neutral utterances were correctly identified, while scared utterances were most frequently misidentified as neutral.

The primary objective of this study was to investigate whether sensitivity to dynamic sensitivity predicts individual

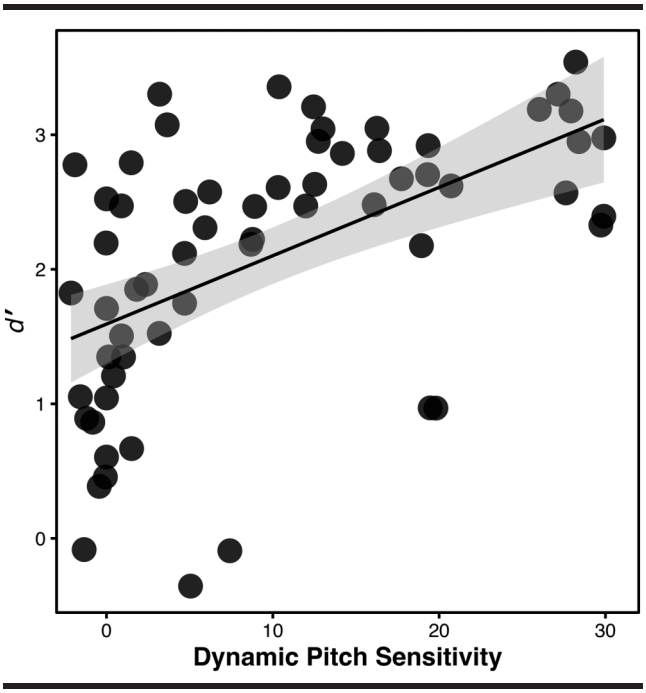
Figure 3. Confusion matrices showing row-normalized vocal emotion recognition by talker sex. Panels display female and male conditions. The top axis lists perceived emotion, and the left axis lists presented emotion. Correct identifications appear along the diagonal, while off-diagonal entries represent misidentification.



differences in vocal emotion recognition accuracy in older listeners. A linear mixed-effects model with d' as the dependent variable, dynamic pitch slope and age as fixed effects, and participant as random intercept indicated that dynamic pitch slope was a significant predictor of vocal emotion recognition accuracy (β = 0.02, *SE* = 0.01, 95% CI [0.003, 0.039], $t(58.96) = 2.36$, $p = .02$). Listeners with steeper slopes (indicating strong dynamic pitch sensitivity) achieved higher d' scores (indicating accurate vocal emotion recognition). The positive relationship between dynamic pitch sensitivity and vocal emotion recognition is consistent with Figure 4, which demonstrates that vocal emotion recognition increases with dynamic pitch sensitivity. Furthermore, the relationship remained significant after controlling for age, indicating that sensitivity to dynamic pitch contributes to individual differences in emotion recognition ability independent of age.

The random intercept variance was 0.261 (*SD* = 0.511), which indicates significant between-participants variation in baseline vocal emotion recognition performance. The residual variance was $\sigma^2 = 0.203$ (*SD* = 0.451). The intraclass correlation coefficient (ICC) representing the proportion of total variance attributable to between-participants differences was .56, indicating that approximately 56% of the variance in d' was due to stable individual differences. The model-fit indices were AIC = 126.8 and BIC = 141.5.

Figure 4. Vocal emotion recognition sensitivity (d') as a function of dynamic pitch sensitivity. Each point represents one participant. Linear regression lines with 95% confidence bands are shown separately for each talker condition.



We conducted further analysis to examine whether hearing status (PTA) contributed to additional variance in the model by testing models that included better ear PTA as a fixed factor. We found no improvement to the model fit and therefore removed both right ear PTA (β = −0.012, *SE* = 0.025, 95% CI [−0.061, 0.037], $t(26.03) = -0.49$, $p = .629$) and left ear PTA (β = −0.006, *SE* = 0.024, 95% CI [−0.053, 0.041], $t(26.09) = -0.26$, $p = .800$) from the final model.

Age was a significant negative predictor of vocal emotion recognition (β = −0.05, *SE* = 0.02, 95% CI [−0.079, −0.017], $t(26.02) = -2.97$, $p = .006$). Listeners within the sample achieved lower d' scores, indicating reduced accuracy in recognizing vocal emotions. The decline in accuracy with increasing age is illustrated in Figure 5, demonstrating a consistent decrease in vocal emotion accuracy with increasing age. The current findings replicate established age-related declines in vocal emotion recognition (Baglione et al., 2025; Ruffman et al., 2008).

Discussion

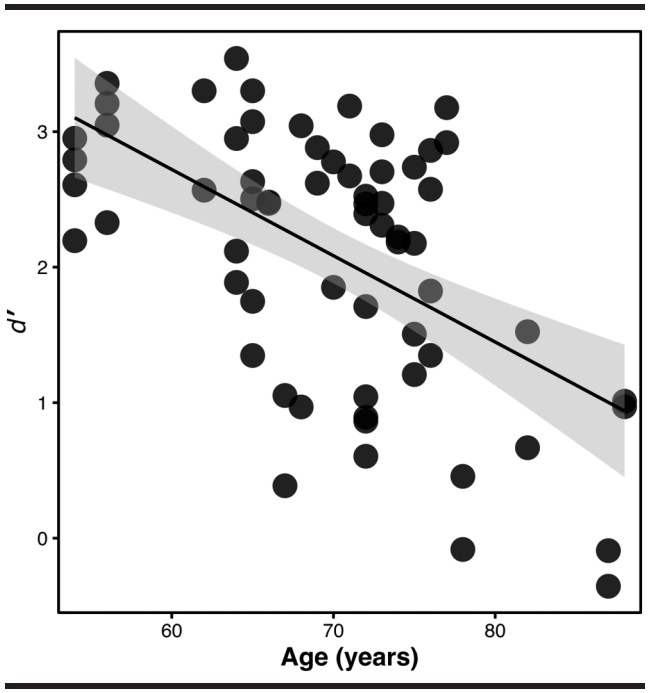
The objective of Study 1 was to examine whether dynamic pitch sensitivity would predict individual differences in vocal emotion recognition. We hypothesized that strong dynamic pitch sensitivity would support accurate emotion recognition, regardless of age. Our findings showed that listeners with greater sensitivity to dynamic pitch achieved higher emotion recognition accuracy.

Dynamic Pitch Sensitivity Predicts Individual Differences in Vocal Emotion Recognition

The findings supported the hypothesis. Dynamic pitch sensitivity significantly predicted vocal emotion recognition. Listeners who demonstrated stronger dynamic pitch sensitivity achieved higher recognition sensitivity (d'). This relationship was still significant after controlling for age and hearing status (PTA) and supports the view that the ability to track dynamic pitch information is functionally linked to the ability to recognize vocal emotions.

This finding is consistent with previous studies examining this relationship in younger adults and mixed-age samples (Dupuis & Pichora-Fuller, 2015; Globerson et al., 2013; Mitchell & Kingston, 2014; Rachman & Başkent, 2025). For instance, Globerson et al. (2013) demonstrated that pitch direction recognition accounted for 31% of the variance in vocal emotion recognition among young adults. Dupuis and Pichora-Fuller (2015) found that *F*0DLs correlated with emotion recognition accuracy in a mixed-age sample, and Mitchell and Kingston (2014) reported that pitch perception showed the strongest correlation with prosodic

Figure 5. Age-related effects on vocal emotion recognition. Each point represents one participant. Regression lines with 95% confidence bands are plotted separately for each talker condition.



emotion comprehension among several auditory measures. More recently, Rachman and Başkent (2025) found that *F*0 sensitivity predicted vocal emotion recognition after age was controlled. The present findings align with this body of work in demonstrating a positive association between dynamic pitch sensitivity and emotion recognition performance.

The present study extends the literature in the following ways. First, we used a dynamic measure of pitch sensitivity (e.g., dynamic pitch identification) instead of static *F*0 measures (e.g., *F*0DLs) used in previous studies (Dupuis & Pichora-Fuller, 2015; Rachman & Başkent, 2025). Since natural emotional prosody is conveyed through dynamic, time-varying pitch contours, static measures may not fully capture the dynamic pitch tracking abilities most relevant to real-world emotion recognition. This study used a dynamic pitch identification task requiring listeners to track the direction of pitch movement over time. This task may more closely resemble the dynamic pitch-tracking demands involved in recognizing vocal emotions in natural speech.

Second, the most direct evidence linking dynamic pitch sensitivity to emotion recognition came from studies of young adults (Globerson et al., 2013). It was not clear whether this relationship also persists in older listeners, who mostly show declines in both abilities. This finding showed that the relationship between dynamic pitch sensitivity and emotion recognition observed in young listeners remains functionally relevant in older listeners. This finding implies that the mechanisms involved in this relationship are preserved across the life span, even as both abilities decline.

Lastly, past studies examining this relationship typically used group comparison designs. Such designs involved either comparing older listeners' performances to those of younger listeners or examining general patterns of aging and their effect on emotion recognition (Dupuis & Pichora-Fuller, 2015; Rachman & Başkent, 2025). These studies did not consider the individual variability that exists among older listeners as a population. Earlier, we also proposed that dynamic pitch sensitivity may explain the individual differences in vocal emotion recognition in older listeners. Here, we recruited only older adults across a wide age range and examined within-group predictors. The findings demonstrate that dynamic pitch sensitivity may explain why some older listeners maintain accurate vocal emotion recognition, whereas others have poor emotion recognition. This finding could be explained in several ways. The significant relationship between dynamic pitch sensitivity and vocal emotion recognition suggests that accurate encoding and extraction of *F*0-based information supports the use of pitch cues in emotion recognition (Mitchell & Kingston, 2014). Because dynamic pitch cues are important carriers of emotional meaning in speech (Banse & Scherer, 1996; Juslin & Laukka, 2003), listeners who can reliably track their movement over time are better equipped to use them to interpret a speaker's emotions. This relationship remained significant after controlling for age and hearing loss, indicating its robustness against age-related effects on both abilities.

Age Predicts Vocal Emotion Recognition

Consistent with our expectations, age was a significant negative predictor of vocal emotion recognition accuracy ($\beta = -0.048$, $SE = 0.015$, 95% CI [−0.078, −0.018], $p = .003$). As age increased, d' also decreased, suggesting age-related decline in performance across the age range of the sample. Even after accounting for dynamic pitch sensitivity, older listeners achieved lower emotion recognition accuracy. This finding replicates studies focused on age-related declines in vocal emotion recognition literature (Baglione et al., 2025; Cannon & Chatterjee, 2022; Christensen et al., 2019; Fan et al., 2024) and extends the literature by showing that age-related decline is observable even within an exclusively older adult cohort.

Although dynamic pitch sensitivity was a significant predictor of the relationship, the remaining between-listeners difference is substantial and unexplained. The remaining difference after accounting for dynamic pitch sensitivity and age indicates that other factors may contribute to individual differences in emotion recognition ability. It is possible that the remaining contribution stems from aspects of auditory processing not represented by the

dynamic pitch identification task (e.g., temporal resolution, spectral processing). Other contributing factors may include cognitive factors such as working memory and processing speed (Pichora-Fuller et al., 2016; Wingfield et al., 2005), shifts in cue weighting (Ben-David et al., 2019; Dor et al., 2022; Lin et al., 2024), and shifts in motivation (Carstensen, 1995; Carstensen et al., 1999). The unexplained between-listeners difference suggests that vocal emotion recognition is supported by multiple processes that may decline with age.

Individual Variability

Consistent with previous studies, we observed large interindividual variability in both dynamic pitch sensitivity and vocal emotion recognition. Psychometric functions for pitch direction identification revealed steep slopes, which indicate accurate tracking of dynamic pitch, while others showed shallow slopes representing poor dynamic pitch tracking. This variability replicates findings reported by Jeon and Heinrich (2022), Shen et al. (2016), and Souza et al. (2011). Souza et al. reported that older listeners showed variability in their ability to perceive dynamic pitch. Shen et al. replicated the study in a new sample of older listeners while strictly controlling for tonal language and musical experience. They found similar variability in dynamic pitch sensitivity within their sample. Their results also confirm the findings of a similar study examining pitch perception in older listeners (Jeon & Heinrich, 2022). In this study, some older listeners showed strong sensitivity to pitch changes, while others required large pitch variations to recognize dynamic pitch differences. Similar variability was observed in vocal emotion recognition. The analysis revealed an ICC of .56, implying that about 56% of the differences in emotion recognition were attributable to differences between listeners. Some listeners consistently recognized emotions accurately, while others also consistently performed poorly. Several factors may account for the variability in both abilities. A decline in periodicity encoding due to aging may weaken access to acoustic information required for accurate dynamic pitch perception (Clinard et al., 2010; Souza et al., 2011). Suprathreshold deficits in temporal processing and frequency selectivity (Füllgrabe et al., 2015; Shen & Souza, 2017) may also degrade the precise encoding of *F*0 contours.

Study 2

In Study 2, we asked whether dynamic pitch sensitivity predicts individual differences in vocal emotion recognition in listening conditions degraded by surrounding noise. We hypothesized that stronger dynamic pitch sensitivity would be associated with accurate vocal emotion recognition, regardless of age and listening condition. We also assessed how closely subjective self-reports of emotions in everyday situations corresponded to objective measures of listeners' ability to recognize vocal emotions.

Method

Participants

Twenty-one listeners (13 female, 8 male) aged 56–88 years (M = 74.24, SD = 8.72) who had participated in Study 1 returned to participate in this study, ensuring that they had already completed the dynamic pitch identification and vocal emotion tasks in quiet. We used the same inclusion criteria as in Study 1. Listeners were included if the better-ear four-frequency PTA was between 0 and 55 dB HL. Mean right ear PTA was 28.57 dB HL, and mean left ear PTA was 28.73 dB HL. Figure 6 shows the audiometric thresholds of listeners included in Study 2. Because Study 2 focused on vocal emotion recognition in noise, the Quick Speech-in-Noise Test (Killion et al., 2004) and the Acceptable Noise Level Test (Nabelek et al., 1991) were included to better characterize participants' speech-in-noise abilities and listening comfort. These measures were included for descriptive purposes only and were not included in the analyses as predictors. The participant demographics are presented in Table 2.

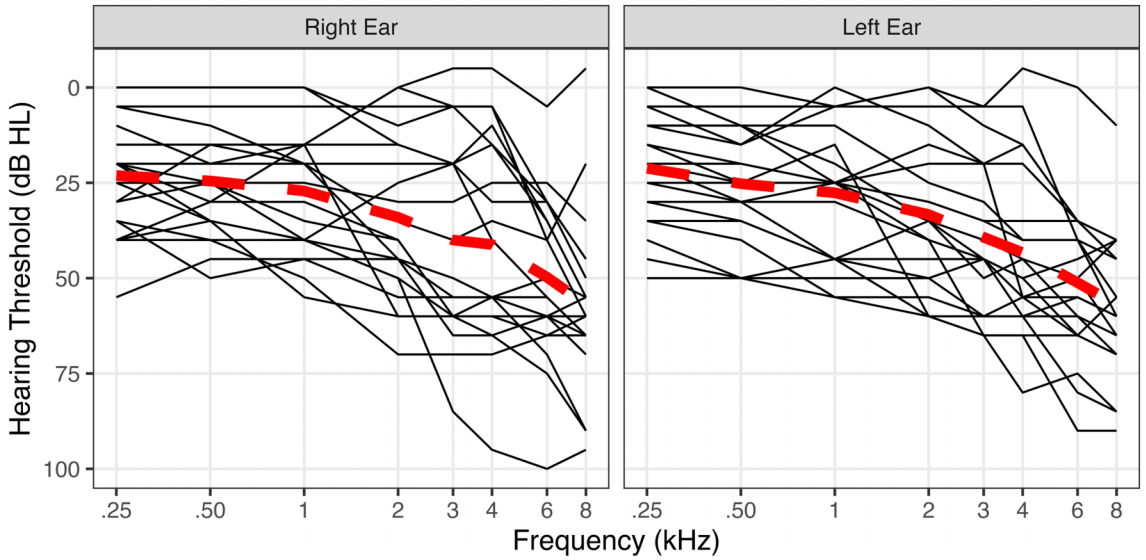
Stimuli

The target sentences were similar to the vocal emotion stimuli used in Study 1. Each sentence was semantically neutral and produced by one male and one female talker conveying "happy, sad, angry, scared, or neutral" prosody. In Study 2, these sentences were presented in a surrounding multitalker babble to mirror a common real-world listening scenario. The babble materials were obtained from the ALLSSTAR corpus (Bradlow, n.d.) and consisted of running speech narratives produced by primary English speakers. The masker consisted of three independent male–female talker pairs, each of which was presented from three separate loudspeaker locations (described more fully below).

Questionnaire

The 16-item EMO-CHeQ (Singh et al., 2019), which assesses self-perceived difficulty in recognizing and conveying emotions in speech, was used. Each statement (e.g., "I have difficulty identifying the emotions expressed by people I interact with on a regular basis," "It is harder for me to identify emotions when I am in a noisy environment") was rated on a 5-point Likert scale from 1 (*strongly disagree*) to

Figure 6. Right and left ear audiograms from all participants (black lines) in Study 2. The red dashed lines indicate the overall mean threshold.



5 (*strongly agree*), with higher ratings indicating greater perceived difficulty.

Test Room Setup

In real-world listening environments, target and competing sources are rarely co-located. Spatial separation of target and masker is relevant to the present study for two reasons. First, it mirrors naturalistic listening conditions; listeners routinely encounter situations in which they must attend to one spatially located talker while ignoring others at different positions. Second, spatial separation may provide a form of degradation that is challenging but potentially recoverable, allowing the examination of emotion recognition under conditions in which listeners have access to cues that could support segregation of the target voice. Accordingly, listeners were tested in a sound-attenuated room containing four loudspeakers placed 90° from each other. Each loudspeaker was positioned 1 m away from the listener.

Table 2. Participant demographics for Study 2.

Characteristic	*M*	*SD*
Left ear PTA (dB HL)	28.73	16.06
Right ear PTA (dB HL)	28.57	15.86
Age (years)	74.24	8.72
MoCA score	27.33	1.43
QuickSIN (dB)	4.05	3.12
ANL	3.29	3.13

Note. PTA = pure-tone average; MoCA = Montreal Cognitive Assessment; QuickSIN = Quick Speech-in-Noise Test; ANL = Acceptable Noise Level Test.

Vocal emotion stimuli were delivered from a front loudspeaker (Yamaha HS50M) at 0° azimuth. Two-talker babble was presented from each of the other three loudspeakers (Gallo Acoustics A'Diva Bass Ball loudspeakers) at 90°, 180°, and 270° azimuths, each calibrated to the same overall level, resulting in a six-talker babble. Two signal-to-noise ratios (SNRs), 3 dB (difficult) and 10 dB (easier), were selected to represent challenging and favorable everyday listening conditions (Smeds et al., 2015; Wu et al., 2018). The listeners were seated at the center of the array and remained in that position throughout the test. Each participant's audiogram was used to compute a listener-specific presentation level, following the procedure described in Study 1, to ensure consistent listener-specific audibility across participants. These levels were then applied to the vocal emotion recognition in noise task so that the vocal emotion stimuli and background babble maintained the 3 and 10 dB SNRs as intended. All stimulus presentations and response logging were controlled using MATLAB (The MathWorks Inc., 2024).

Procedure

The testing was performed in a single session. The EMO-CHeQ questionnaire was first administered to assess self-reported difficulty in recognizing emotions in adverse listening conditions. All questionnaire items were completed. After that, listeners were seated in a sound-attenuated room for the vocal emotion recognition task. The vocal emotion recognition task was administered in the same way as in Study 1. The order of the talker presentation and SNR was counterbalanced across listeners. Participants' responses were

recorded as overall percent correct scores and confusion matrices. All the tests were conducted unaided.

Statistical Analysis

Emotion recognition accuracy was quantified using d', as described in Study 1. The d' values were computed separately for each emotion category from the confusion matrices and then averaged to yield a composite measure of emotion recognition sensitivity for each participant, speaker, and SNR condition. This resulted in four d' values per participant (2 speakers × 2 SNR levels), which were entered into the mixed-effects model with SNR as a categorical predictor.

All analyses were conducted in R (Version 4.4.1) using the lme4 and lmerTest packages, with the same procedures for d' calculation and model diagnostics as described for Study 1. Ninety-five percent CIs were estimated using parametric bootstrapping with 5,000 resamples. The primary outcome of measure was sensitivity (d'), which was calculated from the confusion matrices. Prior to modeling, continuous predictors were mean-centered. Age was centered at the sample mean. Hearing status was indexed by PTA, computed as the mean of the left and right PTA and then centered. Sensitivity to dynamic pitch was indexed by the slope of the psychometric function for pitch direction identification, averaged across the two base frequencies and mean-centered. SNR was treated as a categorical factor with 10 dB SNR as the reference level.

Linear mixed-effects models were used to examine whether sensitivity to dynamic pitch predicted emotion recognition sensitivity (d') under masking conditions while accounting for individual differences in age and hearing status (PTA). Participant was included as a random intercept to account for repeated observations across SNR conditions and speakers. Model building proceeded sequentially: A baseline random-intercept-only model was first established, and then fixed effects were added incrementally (dynamic pitch slope, SNR, Dynamic Pitch Slope × SNR interaction, age, and PTA). Models were estimated using maximum likelihood, and comparisons were made using likelihood ratio tests and AIC and BIC values. Attempts to include random slopes resulted in convergence failures; therefore, these terms were not retained. Inspection of residual histograms and Q–Q plots indicated that the assumptions of normality were met. Pairwise contrasts between SNR conditions were estimated with emmeans using Kenward–Roger degrees of freedom, which provides accurate p values and CIs in small samples (Kenward & Roger, 1997). The final model was specified as

$$d' \sim \text{Dynamic Pitch Slope} * \text{SNR} + \text{Age} + \text{PTA} + (1 \mid \text{Participant}) \quad (3)$$

In addition to mixed-effects models, we examined the relationship between subjective and objective measures of emotional communication. Pearson's correlations were used to assess the association between EMO-CHeQ scores and performance on the emotion recognition task. Responses to the EMO-CHeQ were rated on a 5-point Likert scale ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). Item scores were averaged to produce a single value ranging from 1 to 5, with higher scores indicating greater perceived difficulty recognizing emotional cues.

Results

Descriptive Statistics

Data from 21 listeners were included in the analyses. Mean emotion recognition sensitivity was comparable across SNR conditions (10 dB: $M = 2.89$, $SE = 0.12$; 3 dB: $M = 2.84$, $SE = 0.12$). Descriptive statistics for Study 2 participants are presented in Table 2.

Vocal Emotion Recognition

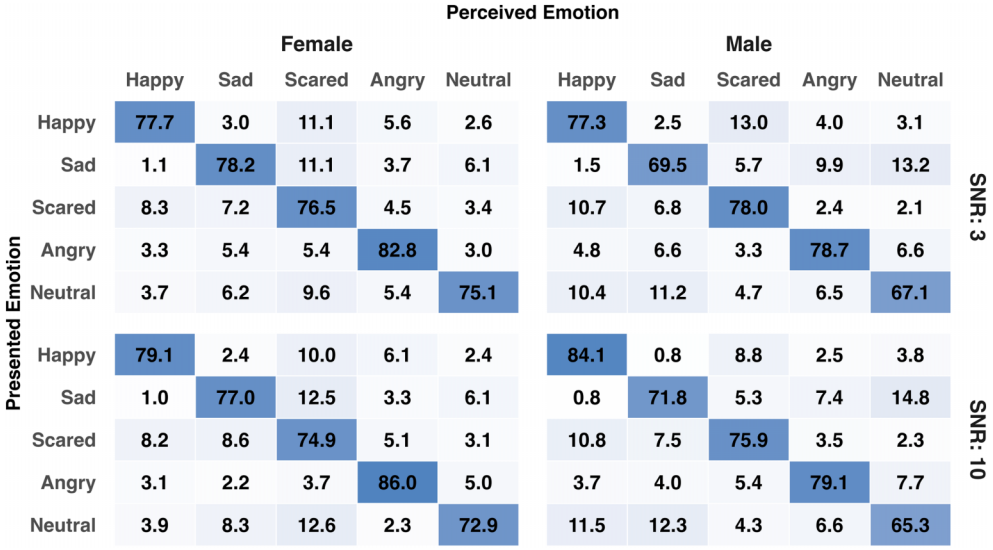
Overall Emotion Recognition Accuracy

Figure 7 provides an overview of the general response patterns, displaying percentage-based confusion matrices across all listeners. Each quadrant corresponds to one talker (female or male) and one SNR condition (10 or 3 dB). Overall accuracy was similar at 10 and 3 dB SNR. Emotion recognition was generally accurate across conditions, with correct identifications ranging from 65.3 to 86.0. Female talker was identified more accurately than male talker.

The primary objective was to determine whether sensitivity to dynamic pitch supports individual differences in vocal emotion recognition in listening conditions degraded by surrounding noise. Dynamic pitch slope showed a positive but nonsignificant association with emotion recognition sensitivity ($\beta = 0.02$, $SE = 0.01$, 95% CI [−0.004, 0.035], $t(31.2) = 1.54$, $p = .14$). Although the direction of this effect was consistent with Study 1 such that listeners with greater sensitivity to dynamic pitch achieved higher d' scores, this relationship did not reach statistical significance. Figure 8 illustrates the trend toward a higher d' among listeners with steeper dynamic pitch slopes.

As a contextual analysis, we examined the interaction between dynamic pitch slope and SNR conditions. The interaction was not significant ($\beta = 0.00$, $SE = 0.01$, 95% CI [−0.016, 0.018], $t(399) = 0.08$, $p = .940$), indicating that the relationship between dynamic pitch sensitivity and emotion recognition did not differ between the 10 and 3 dB SNR conditions. Similarly, the main effect of SNR was not significant ($\beta = -0.05$, $SE = 0.10$, 95% CI [−0.255, 0.153], $t(399) = -0.46$, $p = .65$). Estimated marginal means were

Figure 7. Confusion matrices showing row-normalized vocal emotion recognition by talker sex and signal-to-noise ratio (SNR). Panels display female and male talkers and 10 and 3 dB SNR conditions. The top axis lists perceived emotion, and the left axis lists presented emotion. Correct identifications appear along the diagonal, while off-diagonal entries represent misidentifications.



equivalent across noise levels (10 dB: $M = 2.89$, $SE = 0.13$; 3 dB: $M = 2.84$, $SE = 0.13$; $p = .647$), suggesting that emotion recognition performance was comparable across noise levels after controlling for individual differences. The random intercept variance was 0.155 ($SD = 0.39$), suggesting meaningful between-participants variability in baseline emotion recognition performance. The residual variance was 1.113 ($SD = 1.05$). The ICC was .12, indicating that approximately 12% of the total variance in d' was attributable to stable between-participants differences. Model fit indices were AIC = 1280.8 and BIC = 1313.1.

Age was a significant predictor of vocal emotion ($\beta = -0.041$, $SE = 0.014$, 95% CI [−0.068, −0.014], $t(21.0) = -2.96$, $p = .008$). Listeners achieved lower d' scores, indicating reduced accuracy in recognizing vocal emotions in the presence of noise. Figure 9 plots d' against age, showing clear downward slopes for SNR conditions, consistent with

Figure 8. Vocal emotion recognition sensitivity (d') as a function of dynamic pitch sensitivity at 3 and 10 dB SNR conditions. Regression lines with 95% confidence bands illustrate the relationship for each talker and noise level.

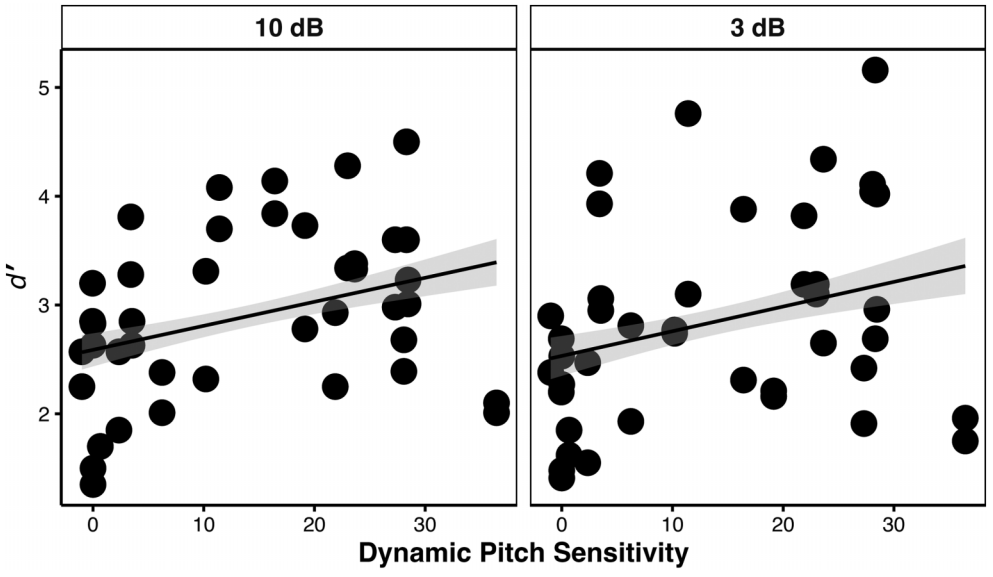
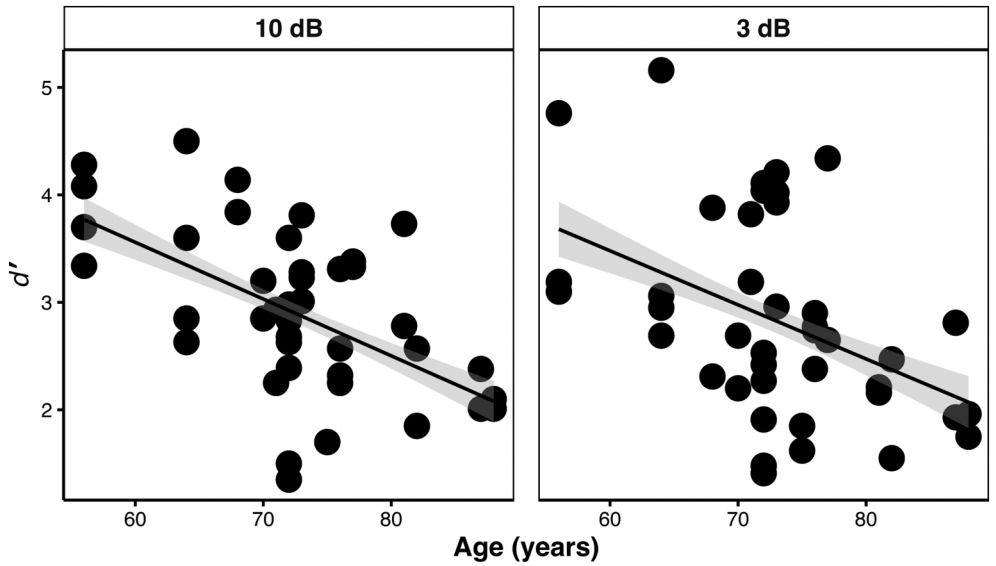


Figure 9. Age effects on emotion recognition at 10 and 3 dB SNR conditions. Points represent individual participants. Lines are linear fits with 95% confidence bands shown separately for each noise condition.



an age-related decline in sensitivity. This finding extends the age effects observed in Study 1 to adverse listening conditions. PTA showed a negative but nonsignificant association with emotion recognition sensitivity ($\beta = -0.014$, $SE = 0.008$, 95% CI [−0.031, 0.002], $t(21.0) = -1.70$, $p = .105$), suggesting that hearing status (PTA) did not explain unique variance beyond age and dynamic pitch sensitivity.

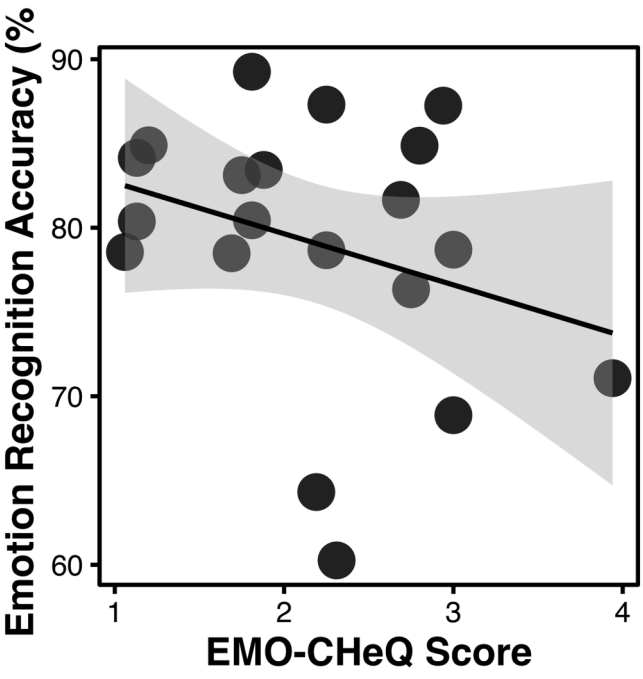
A final secondary objective was to examine whether objective emotion recognition performance corresponded with subjective reports of difficulty recognizing vocal emotions. EMO-CHeQ scores ranged from 1 to 5, with higher scores indicating greater perceived difficulty recognizing emotional cues. Pearson correlation showed a significant negative association between EMO-CHeQ scores and vocal emotion recognition accuracy, $r(82) = -.30$, $p = .005$, 95% CI [−.49, −.10]. Listeners who reported greater ease with emotional communication performed more accurately on the objective task, indicating that subjective reports from the EMO-CHeQ aligned with objective performance. Figure 10 illustrates this relationship, with accuracy declining as the perceived difficulty increased.

Discussion

The primary objective of Study 2 was to examine whether dynamic pitch sensitivity supports individual differences in the recognition of vocal emotions masked by surrounding noise. Having established in Study 1 that dynamic pitch sensitivity predicts individual differences in vocal emotion recognition in quiet, Study 2 assessed whether this relationship is preserved or declines when the listening condition becomes challenging.

Dynamic pitch sensitivity was positively associated with vocal emotion recognition in the noise condition. The

Figure 10. Relationship between Emotional Communication in Hearing Questionnaire (EMO-CHeQ) score and vocal emotion recognition accuracy. Each point represents each listener's accuracy score, along with a regression line and a 95% confidence band.



relationship showed a positive trend consistent with theoretical expectations; listeners with better dynamic pitch sensitivity were better at recognizing vocal emotion. However, this relationship was not statistically significant (p = .135). The 95% CI [−0.004, 0.035] includes zero; as such, the present data cannot rule out either a small positive effect or no effect. Recall that we reasoned that if pitch cues are relatively robust to masking (Guo et al., 2022; Kong & Zeng, 2006), the ability to successfully extract these cues should determine who may have preserved or degraded vocal emotion recognition in noise. The data showed a positive association between dynamic pitch sensitivity and emotion recognition in noise, consistent with this hypothesis. The finding can be interpreted from two complementary perspectives. On one hand, Buss and Bosen (2021) demonstrated that the low frequency bands, which carry *F*0 information, are important for speech recognition in babble. If *F*0 tracking contributes to auditory stream segregation and enables glimpsing (Bregman, 1994; Cooke, 2006), it is reasonable to expect that sensitivity to dynamic pitch cues would support emotion recognition in babble. This acoustic–perceptual mapping is consistent with the positive direction of the data patterns in noise, as well as the similar effect sizes in quiet and noise. On the other hand, babble may mask dynamic pitch cues and reduce reliance on them for emotion recognition. According to Dor et al. (2024), babble increases cognitive load, compelling listeners to consider alternative listening strategies, including reliance on complementary prosodic cues. In such situations, older listeners may weight semantic content more to infer emotion (Ben-David et al., 2019; Dor et al., 2022). These two perspectives are not mutually exclusive. The consistent effect sizes in both studies suggest that dynamic pitch sensitivity may still support vocal emotion recognition in noise. However, its contribution may have been weakened, as reflected by the nonsignificant result, possibly because listeners perhaps supplemented dynamic pitch-based processing with alternative strategies under adverse listening conditions.

Several factors could explain the nonsignificant finding. First, the modest sample size (n = 21) resulted in a larger standard error (SE = 0.010). This may have produced a lot of uncertainty around the estimate and reduced the probability of achieving statistical significance. Second, the variance component suggests that most of the differences in vocal emotion recognition scores were due to trial-to-trial fluctuations rather than stable interindividual variability. Interindividual variance was quite small (τ_{00} = 0.155) compared to the residual variance (σ^2 = 1.113), resulting in a low ICC (.12). It is possible that background noise introduced large trial-to-trial fluctuations, which perhaps limited variance for the listener-level effects (i.e., dynamic pitch sensitivity) to explain. Together, these factors may have contributed to the reduced precision of the estimated effect and increased the likelihood of a nonsignificant result. It is, therefore, important that the nonsignificant is not interpreted as an absence of a relationship between the two abilities. A more accurate interpretation may be that the data points to a small positive effect, but the modest sample size creates considerable uncertainty.

Age was a significant negative predictor of vocal emotion recognition in noise. As age increased, d' also decreased, suggesting age-related decline in performance across the age range of this cohort. This finding extends the age-related declines in emotion recognition observed in Study 1 to adverse listening conditions.

The absence of a significant main effect (β = −0.047) of SNR indicates that emotion recognition performance was comparable at both noise conditions. The 95% CI [−0.004, 0.035] indicates that there is no evidence of a significant difference between 10 and 3 dB SNR conditions. This interpretation is further supported by the marginal means, 10 dB SNR (d' = 2.89) and 3 dB SNR (d' = 2.84). An explanation for this result may be that trial-to-trial variability accounted for most of the individual differences in emotion recognition performance in noise, rather than stable condition-level effects (i.e., SNRs). As a result, the sensitivity to detect the difference between conditions may have been reduced.

We also assessed self-perceived difficulty in recognizing emotions in speech using the EMO-CHeQ. EMO-CHeQ scores were significantly correlated with vocal emotion recognition accuracy. Listeners who reported more difficulty with vocal emotion recognition in everyday situations also performed poorly on the vocal emotion recognition task. This agreement between subjective and objective measures of vocal emotion recognition supports the real-world relevance of the objective emotion recognition task. In other words, the emotion recognition abilities measured under controlled conditions are related to older listeners' perceived experiences of emotional communication in daily life. This finding is also in accordance with previous studies (Goy et al., 2018; Singh et al., 2019). Our finding extends the literature by showing that objective vocal emotion recognition may align with real-world listening experiences. This result also suggests that older listeners may be aware of their limitations in recognizing vocal emotions and that subjective difficulties may mirror objective deficits. A potential clinical implication of this finding is the utility of the EMO-CHeQ as a screening tool for identifying listeners who might benefit from interventions targeting vocal emotion recognition abilities. At the same time, the modest size of the correlation (r = −.30) suggests that other factors influence how objective vocal emotion recognition ability translates into real-world function. In real-world emotional communication, listeners may access and benefit from visual cues, talker familiarity, semantic contexts, the Lombard effect, and/or opportunities for conversation repair (Clark & Schaefer, 1989; Dupuis &

Pichora-Fuller, 2015; Lu & Cooke, 2008; Paulmann & Pell, 2011)—all advantages unavailable in objective tasks. In addition, individual differences in compensatory strategies and self-awareness (Bally, 2003; Kricos, 2000) may also influence the agreement between objective and subjective measures.

General Discussion

The present study examined whether sensitivity to dynamic pitch predicts individual differences in vocal emotion recognition and whether this relationship extends to listening conditions characteristic of everyday listening environments. Across the two studies, the findings revealed a pattern: Dynamic pitch sensitivity showed a positive association with vocal emotion, with similar effect sizes across quiet and background noise. In quiet, this relationship was statistically significant; in noise, it was not.

Comparison of Primary Findings

A noteworthy finding in this study is the consistent effect sizes across both studies (Study 1: $\beta = 0.02$; Study 2: $\beta = 0.02$), despite differences in statistical significance ($p = .012$ vs. $p = .135$). The inconsistency may be best explained by methodological factors. For instance, Study 2 had a modest sample size compared to Study 1. This resulted in a larger standard error ($SE = 0.010$ vs. 0.006) and a wider CI that includes zero. These factors may have collectively contributed to reducing the precision of the estimated effect of Study 2, not the magnitude of the effect itself.

The consistency in effect sizes across quiet and noise conditions is an important contribution to the vocal emotion literature. Previous studies established that pitch sensitivity is associated with vocal emotion recognition, but most of these studies were conducted in quiet (Dupuis & Pichora-Fuller, 2015; Globerson et al., 2013; Mitchell & Kingston, 2014; Rachman & Başkent, 2025). It was unclear whether this relationship would be maintained under adverse listening conditions. This environment is important because it has practical relevance for everyday communication. The current findings indicate that although the relationship was statistically detected only in quiet, comparable effect sizes across conditions suggest that the contribution of dynamic pitch sensitivity may persist across listening conditions, with increased trial-level variability in noise likely weakening the ability to detect it statistically. Collectively, the findings contribute to the literature by showing that the recognition of vocal emotion relies on sensitivity to dynamic changes in pitch; however, this relationship may be influenced by listening context.

Comparison of variance structures across both studies may also provide further insight into why the relationship between dynamic pitch sensitivity and vocal emotion recognition was clear in quiet but unclear under masking. In Study 1, over half of the variability in emotion recognition was due to stable differences between listeners (ICC = .56), suggesting that listeners were quite consistent across trials. Because of this stability, dynamic pitch significantly predicted vocal emotion recognition performance. However, in Study 2, only 12% (ICC = .12) of the variability was attributed to stable differences between listeners, while the rest of 88% were due to trial-to-trial fluctuations. This shift was due to a large increase in residual variance ($\sigma^2 = 1.113$ vs. 0.203), while between-listeners variance remained relatively similar to Study 1 ($\tau_{00} = 0.155$ vs. 0.261). In other words, individual differences in dynamic pitch sensitivity may still be present but may have been overlaid by elevated trial-to-trial variability due to noise, making it difficult to statistically detect the relationship between the two abilities. The glimpsing hypothesis (Cooke, 2006) may provide some insight into this pattern. In noise, performance may have been driven by access to dynamic pitch cues during momentary masker dips. These moments are mostly unpredictable and independent of the listener's ability. Hence, even a listener with strong dynamic pitch sensitivity may access pitch cues inconsistently across trials.

Underlying Mechanism

These findings also support a hierarchical framework linking *F*0 encoding to functional emotion communication. At the most fundamental level, sensitivity to small differences in *F*0 supports the sensitivity to dynamic pitch (Chatterjee & Peng, 2008; Peng et al., 2009; Souza et al., 2011). The present study extends this hierarchy by showing that dynamic pitch sensitivity, in turn, predicts vocal emotion recognition. This finding establishes a processing chain, from *F*0 discrimination to dynamic pitch sensitivity to vocal emotion recognition, in which each level depends on the integrity of the preceding level. In other words, when sensitivity to small *F*0 differences is degraded, whether due to age-related changes or masking, it may compromise dynamic pitch sensitivity, which in turn may limit the ability to recognize vocal emotions. This sequence, in which deficits at lower levels of auditory processing carry forward to degrade higher level perceptual abilities, is consistent with evidence that difficulties perceiving simple *F*0 changes affect the use of *F*0 in more complex listening tasks (Shen & Souza, 2018; Souza et al., 2011). The main theoretical contribution of this study is its positioning of vocal emotion recognition as the end result of a chain of auditory processes, progressing from *F*0 discrimination to dynamic pitch sensitivity and, ultimately, to vocal emotion recognition, an ability relevant for social cognition.

Age Effect

Age was a consistent negative predictor of vocal emotion recognition accuracy across both studies, even after

accounting for dynamic pitch sensitivity and hearing status (PTA). This age effect indicates that multiple mechanisms apart from dynamic pitch sensitivity contribute to age-related decline in vocal emotion recognition. The findings align with a large body of literature on age-related deficits in emotion recognition (Baglione et al., 2025; Cannon & Chatterjee, 2022; Christensen et al., 2019; Ruffman et al., 2008). The literature proposes several theories to explain these age-related declines in vocal emotion recognition. Declines in suprathreshold hearing sensitivity and auditory processing may contribute to declines in emotion, even though mild-to-moderate elevations in pure-tone thresholds are not predictive (Orbelo et al., 2005; Pichora-Fuller & Souza, 2003). This theory is consistent with our results showing that hearing status (PTA) did not predict vocal emotion in either study. Age-related declines in vocal emotion have also been attributed to cognitive decline (e.g., working memory, executive function, and processing speed; Orbelo et al., 2005) and neuroanatomical changes in emotion processing regions, particularly the right hemisphere (Ruffman et al., 2008; Wright et al., 2018). Lastly, socioemotional selectivity theory proposes that aging leads to shifts in motivation toward positive emotional experiences, giving rise to a "positivity effect" such that older adults prioritize negative emotions less (Carstensen, 1995; Carstensen et al., 1999). This shift in priority may also contribute to age-related declines in emotion recognition. Based on these findings, multiple factors beyond audibility or dynamic pitch sensitivity may affect how older adults respond to talker emotions in everyday communication.

Broader Implications

The findings support a sensory–perceptual account of vocal emotion recognition. The ability to discriminate *F*0 at the sensory level may support dynamic pitch sensitivity, which in turn contributes to high-level prosodic comprehension (i.e., vocal emotion recognition), and this contribution appears stable across listening environments. The consistent age effect, independent of dynamic pitch sensitivity and hearing status (PTA), indicates that multiple mechanisms contribute to the age-related decline in vocal emotion recognition. The primary finding that dynamic pitch sensitivity contributes to individual differences in vocal emotion recognition suggests that interventions focused on improving pitch direction discrimination might benefit emotional communication, although this requires direct testing.

Limitations and Future Directions

Several limitations of this study should be considered when interpreting its findings. The modest sample in Study 2, combined with the large variability introduced by listening in background noise, may have limited power to detect small effects. The similar effect sizes observed in quiet and noise, coupled with the large standard error in noise, suggest that the relationship between dynamic pitch sensitivity and vocal emotion recognition may exist in noise but was difficult to detect due to variability in performance. Replicating this study with larger samples will be needed to examine whether the effect is reliable and the specific conditions under which it persists or breaks down. Although audibility was controlled for each listener, minor differences in effective level across listeners cannot be fully ruled out. A replication of this work could ensure stricter control of sensation levels. Cognitive factors, including working memory, selective attention, and executive control, are known to impact listening in noise. Individual differences in these cognitive factors may contribute to differences in vocal emotion recognition in noise. Without examining these factors, it is impossible to determine whether sensory or cognitive deficits contributed to listeners' performance. Further work is necessary to examine the contributions of cognitive factors to vocal emotion recognition in noise. We also did not identify which alternative cues listeners used when dynamic pitch was uninformative or degraded, so the relative contributions of different prosodic features, such as intensity and duration, cannot be determined. Without this information, the robustness of vocal emotion recognition in noisy conditions cannot be attributed solely to cue redundancy. Finally, although the chosen SNRs are based on real-world SNRs among older adults, the emotion stimuli may not have adequately captured real-world emotional communication. In everyday listening situations, dynamic pitch cues are more likely to be exaggerated, and communication may also be supported by visual cues and contextual information to ease emotional communication. As such, these findings should not be interpreted as a direct estimate of real-world emotional communication.

Conclusions

Sensitivity to dynamic pitch is an important predictor of individual differences in vocal emotion recognition in older listeners. The relationship was significant in quiet but not in noise, although similar effect sizes suggest that individual sensitivity to dynamic pitch may be an important cue in noise. Age-related decline was evident across studies and was not explained by hearing status (PTA) and dynamic pitch sensitivity. The agreement between objective performance and subjective reports supports the functional relevance of these findings.

Data Availability Statement

The data sets generated and analyzed in this study are available from the corresponding author upon reasonable request.

Acknowledgments

This research was supported by a grant from the Alumnae of Northwestern University (to Pamela Souza). The authors are grateful to Aditya M. Kulkarni for his help with the vocal emotion program used in this study, Nathan Fishpaugh for designing the vocal emotion in noise program, Kendra Marks for assistance with participant scheduling and data collection, and Jason Sanchez for helpful recommendations on the study.

References

Adolphs, R. (2002). Neural systems for recognizing emotion. *Current Opinion in Neurobiology, 12*(2), 169–177. https://doi.org/10.1016/s0959-4388(02)00301-x

American National Standards Institute. (2018). *Specification for audiometers* (ANSI S3.6-2018 [R2023]). https://webstore.ansi.org/standards/asa/asaansis32018r2023

Anderson, S., Parbery-Clark, A., Yi, H.-G., & Kraus, N. (2011). A neural basis of speech-in-noise perception in older adults. *Ear and Hearing, 32*(6), Article 750. https://doi.org/10.1097/aud.0b013e31822229d3

Baglione, H., Coulombe, V., Martel-Sauvageau, V., & Monetta, L. (2025). The impacts of aging on the comprehension of affective prosody: A systematic review. *Applied Neuropsychology: Adult, 32*(4), 1189–1204. https://doi.org/10.1080/23279095.2023.2245940

Bally, S. J. (2003). Conversation made easy, speechreading, and conversation strategies training for children with hearing loss. *Ear and Hearing, 24*(1), 96–97. https://doi.org/10.1097/00003446-200302000-00012

Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology, 70*(3), 614–636. https://doi.org/10.1037/0022-3514.70.3.614

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Ben-David, B. M., Gal-Rosenblum, S., van Lieshout, P. H. H. M., & Shakuf, V. (2019). Age-related differences in the perception of emotion in spoken language: The relative roles of prosody and semantics. *Journal of Speech, Language, and Hearing Research, 62*(4S), 1236–1250. https://doi.org/10.1044/2018_JSLHR-H-ASCC7-18-0166

Bradlow, A. R. (n.d.). *ALLSSTAR* (Archive of L1 and L2 Scripted and Spontaneous Transcripts and Recordings) [Data set]. Northwestern University, Speech Communication Laboratory. https://speechbox.linguistics.northwestern.edu/allsstar

Bregman, A. S. (1994). *Auditory scene analysis: The perceptual organization of sound*. MIT Press. https://doi.org/10.1121/1.408434

Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference: Understanding AIC and BIC in model selection. *Sociological Methods & Research, 33*(2), 261–304. https://doi.org/10.1177/0049124104268644

Buss, E., & Bosen, A. (2021). Band importance for speech-in-speech recognition. *JASA Express Letters, 1*(8), Article 084401. https://doi.org/10.1121/10.0005762

Cannon, S. A., & Chatterjee, M. (2019). Voice emotion recognition by children with mild-to-moderate hearing loss. *Ear and Hearing, 40*(3), 477–492. https://doi.org/10.1097/AUD.0000000000000637

Cannon, S. A., & Chatterjee, M. (2022). Age-related changes in voice emotion recognition by post-lingually deafened listeners with cochlear implants. *Ear and Hearing, 43*(2), 323–334. https://doi.org/10.1097/aud.0000000000001095

Carson, N., Leach, L., & Murphy, K. J. (2018). A re-examination of Montreal Cognitive Assessment (MoCA) cutoff scores. *International Journal of Geriatric Psychiatry, 33*(2), 379–388. https://doi.org/10.1002/gps.4756

Carstensen, L. L. (1995). Evidence for a life-span theory of socioemotional selectivity. *Current Directions in Psychological Science, 4*(5), 151–156. https://doi.org/10.1111/1467-8721.ep11512261

Carstensen, L. L., Isaacowitz, D. M., & Charles, S. T. (1999). Taking time seriously: A theory of socioemotional selectivity. *American Psychologist, 54*(3), 165–181. https://doi.org/10.1037/0003-066x.54.3.165

Chatterjee, M., & Peng, S.-C. (2008). Processing *F*0 with cochlear implants: Modulation frequency discrimination and speech intonation recognition. *Hearing Research, 235*(1–2), 143–156. https://doi.org/10.1016/j.heares.2007.11.004

Christensen, J. A., Sis, J., Kulkarni, A. M., & Chatterjee, M. (2019). Effects of age and hearing loss on the recognition of emotions in speech. *Ear and Hearing, 40*(5), 1069–1083. https://doi.org/10.1097/AUD.0000000000000694

Clark, H. H., & Schaefer, E. F. (1989). Contributing to discourse. *Cognitive Science, 13*(2), 259–294. https://doi.org/10.1207/s15516709cog1302_7

Clinard, C. G., Tremblay, K. L., & Krishnan, A. R. (2010). Aging alters the perception and physiological representation of frequency: Evidence from human frequency-following response recordings. *Hearing Research, 264*(1–2), 48–55. https://doi.org/10.1016/j.heares.2009.11.010

Cooke, M. (2006). A glimpsing model of speech perception in noise. *The Journal of the Acoustical Society of America, 119*(3), 1562–1573. https://doi.org/10.1121/1.2166600

Deroche, M. L. D., Kulkarni, A. M., Christensen, J. A., Limb, C. J., & Chatterjee, M. (2016). Deficits in the sensitivity to pitch sweeps by school-aged children wearing cochlear implants. *Frontiers in Neuroscience, 10,* Article 73. https://doi.org/10.3389/fnins.2016.00073

Dor, Y. I., Algom, D., Shakuf, V., & Ben-David, B. M. (2022). Age-related changes in the perception of emotions in speech: Assessing thresholds of prosody and semantics recognition in noise for young and older adults. *Frontiers in Neuroscience, 16,* Article 846117. https://doi.org/10.3389/fnins.2022.846117

Dor, Y. I., Algom, D., Shakuf, V., & Ben-David, B. M. (2024). Age-related differences in processing of emotions in speech disappear with babble noise in the background. *Cognition and Emotion, 38*(5), 1–10. https://doi.org/10.1080/02699931.2024.2351960

Dupuis, K., & Pichora-Fuller, M. K. (2010). Use of affective prosody by young and older adults. *Psychology and Aging, 25*(1), 16–29. https://doi.org/10.1037/a0018777

Dupuis, K., & Pichora-Fuller, M. K. (2015). Aging affects identification of vocal emotions in semantically neutral sentences. *Journal of Speech, Language, and Hearing Research, 58*(3), 1061–1076. https://doi.org/10.1044/2015_JSLHR-H-14-0256

Fan, X., Tang, E., Zhang, M., Lin, Y., Ding, H., & Zhang, Y. (2024). Decline of affective prosody recognition with a positivity bias among older adults: A systematic review and meta-analysis. *Journal of Speech, Language, and Hearing Research, 67*(10), 3862–3879. https://doi.org/10.1044/2024_JSLHR-23-00775

Füllgrabe, C., Moore, B. C., & Stone, M. A. (2015). Age-group differences in speech identification despite matched audiometrically normal hearing: Contributions from auditory temporal

processing and cognition. *Frontiers in Aging Neuroscience, 6,* Article 347. https://doi.org/10.3389/fnagi.2014.00347

Globerson, E., Amir, N., Golan, O., Kishon-Rabin, L., & Lavidor, M. (2013). Psychoacoustic abilities as predictors of vocal emotion recognition. *Attention, Perception, & Psychophysics, 75*(8), 1799–1810. https://doi.org/10.3758/s13414-013-0518-x

Goy, H., Pichora-Fuller, M. K., Singh, G., & Russo, F. A. (2018). Hearing aids benefit recognition of words in emotional speech but not emotion identification. *Trends in Hearing, 22,* Article 2331216518801736. https://doi.org/10.1177/2331216518801736

Green, T., Faulkner, A., & Rosen, S. (2002). Spectral and temporal cues to pitch in noise-excited vocoder simulations of continuous-interleaved-sampling cochlear implants. *The Journal of the Acoustical Society of America, 112*(5), 2155–2164. https://doi.org/10.1121/1.1506688

Guo, T., Zhu, Z., Kidani, S., & Unoki, M. (2022). Contribution of common modulation spectral features to vocal-emotion recognition of noise-vocoded speech in noisy reverberant environments. *Applied Sciences, 12*(19), Article 9979. https://doi.org/10.3390/app12199979

Hautus, M. J., Macmillan, N. A., & Creelman, C. D. (2021). *Detection theory: A user's guide* (3rd ed.). Routledge. https://doi.org/10.4324/9781003203636

Irino, T., Hanatani, Y., Kishida, K., Naito, S., & Kawahara, H. (2024). Effects of age and hearing loss on speech emotion discrimination. *Scientific Reports, 14*(1), Article 18328. https://doi.org/10.1038/s41598-024-69216-7

Jeon, H.-S., & Heinrich, A. (2022). Perceptual asymmetry between pitch peaks and valleys. *Speech Communication, 140,* 109–127. https://doi.org/10.1016/j.specom.2022.04.001

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. https://doi.org/10.1037/0033-2909.129.5.770

Kenward, M. G., & Roger, J. H. (1997). Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics, 53*(3), 983–997. https://doi.org/10.2307/2533558

Killion, M. C., Niquette, P. A., Gudmundsen, G. I., Revit, L. J., & Banerjee, S. (2004). Development of a quick speech-in-noise test for measuring signal-to-noise ratio loss in normal-hearing and hearing-impaired listeners. *The Journal of the Acoustical Society of America, 116*(4), 2395–2405. https://doi.org/10.1121/1.1784440

Kong, Y.-Y., & Zeng, F.-G. (2006). Temporal and spectral cues in Mandarin tone recognition. *The Journal of the Acoustical Society of America, 120*(5), 2830–2840. https://doi.org/10.1121/1.2346009

Kotz, S. A., & Paulmann, S. (2011). Emotion, language, and the brain. *Language and Linguistics Compass, 5*(3), 108–125. https://doi.org/10.1111/j.1749-818x.2010.00267.x

Kricos, P. B. (2000). The influence of nonaudiological variables on audiological rehabilitation outcomes. *Ear and Hearing, 21*(4), 7S–14S. https://doi.org/10.1097/00003446-200008001-00003

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software, 82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Lieberman, P., & Michaels, S. B. (1962). Some aspects of fundamental frequency and envelope amplitude as related to the emotional content of speech. *The Journal of the Acoustical Society of America, 34*(7), 922–927. https://doi.org/10.1121/1.1918222

Lin, Y., Ye, X., Zhang, H., Xu, F., Zhang, J., Ding, H., & Zhang, Y. (2024). Category-sensitive age-related shifts between prosodic and semantic dominance in emotion perception linked to cognitive capacities. *Journal of Speech, Language, and Hearing Research, 67*(12), 4829–4849. https://doi.org/10.1044/2024_JSLHR-23-00817

Lu, Y., & Cooke, M. (2008). Speech production modifications produced by competing talkers, babble, and stationary noise. *The Journal of the Acoustical Society of America, 124*(5), 3261–3275. https://doi.org/10.1121/1.2990705

Macmillan, N. A., & Kaplan, H. L. (1985). Detection theory analysis of group data: Estimating sensitivity from average hit and false-alarm rates. *Psychological Bulletin, 98*(1), 185–199. https://doi.org/10.1037/0033-2909.98.1.185

Mattys, S. L., Davis, M. H., Bradlow, A. R., & Scott, S. K. (2012). Speech recognition in adverse conditions: A review. *Language and Cognitive Processes, 27*(7–8), 953–978. https://doi.org/10.1080/01690965.2012.705006

Mitchell, R. L., & Kingston, R. A. (2014). Age-related decline in emotional prosody discrimination: Acoustic correlates. *Experimental Psychology, 61*(3), 215–223. https://doi.org/10.1027/1618-3169/a000241

Morgan, S. D. (2021). Comparing emotion recognition and word recognition in background noise. *Journal of Speech, Language, and Hearing Research, 64*(5), 1654–1665. https://doi.org/10.1044/2021_JSLHR-20-00153

Nabelek, A. K., Tucker, F. M., & Letowski, T. R. (1991). Toleration of background noises: Relationship with patterns of hearing aid use by elderly persons. *Journal of Speech and Hearing Research, 34*(3), 679–685. https://doi.org/10.1044/jshr.3403.679

Nasreddine, Z. S., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., Cummings, J. L., & Chertkow, H. (2005). The Montreal Cognitive Assessment, MoCA: A brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society, 53*(4), 695–699. https://doi.org/10.1111/j.1532-5415.2005.53221.x

Nilsson, M., Soli, S. D., & Sullivan, J. A. (1994). Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise. *The Journal of the Acoustical Society of America, 95*(2), 1085–1099. https://doi.org/10.1121/1.408469

Nussbaum, C., Schirmer, A., & Schweinberger, S. R. (2023). Electrophysiological correlates of vocal emotional processing in musicians and non-musicians. *Brain Sciences, 13*(11), Article 1563. https://doi.org/10.3390/brainsci13111563

Orbelo, D. M., Grim, M. A., Talbott, R. E., & Ross, E. D. (2005). Impaired comprehension of affective prosody in elderly subjects is not predicted by age-related hearing loss or age-related cognitive decline. *Journal of Geriatric Psychiatry and Neurology, 18*(1), 25–32. https://doi.org/10.1177/0891988704272214

Paulmann, S., & Pell, M. D. (2011). Is there an advantage for recognizing multi-modal emotional stimuli? *Motivation and Emotion, 35*(2), 192–201. https://doi.org/10.1007/s11031-011-9206-0

Paulmann, S., Pell, M. D., & Kotz, S. A. (2008). How aging affects the recognition of emotional speech. *Brain and Language, 104*(3), 262–269. https://doi.org/10.1016/j.bandl.2007.03.002

Pawełczyk, A., Łojek, E., Radek, M., & Pawełczyk, T. (2021). Prosodic deficits and interpersonal difficulties in patients with schizophrenia. *Psychiatry Research, 306,* Article 114244. https://doi.org/10.1016/j.psychres.2021.114244

Peng, S.-C., Lu, N., & Chatterjee, M. (2009). Effects of cooperating and conflicting cues on speech intonation recognition by cochlear implant users and normal hearing listeners. *Audiology and Neurotology, 14*(5), 327–337. https://doi.org/10.1159/000212112

Pichora-Fuller, M. K., Kramer, S. E., Eckert, M. A., Edwards, B., Hornsby, B. W., Humes, L. E., Lemke, U., Lunner, T.,

Matthen, M., & Mackersie, C. L. (2016). Hearing impairment and cognitive energy: The Framework for Understanding Effortful Listening (FUEL). *Ear and Hearing, 37*(Suppl. 1), 5S–27S. https://doi.org/10.1097/aud.0000000000000312

Pichora-Fuller, M. K., & Souza, P. E. (2003). Effects of aging on auditory processing of speech. *International Journal of Audiology, 42*(Suppl. 2), 11–16. https://doi.org/10.3109/14992020309074638

Rachman, L., & Başkent, D. (2025). Characterization of voice cue sensitivity and vocal emotion recognition across the adult lifespan. *Proceedings of Interspeech, 2025,* 1298–1302. https://doi.org/10.21437/Interspeech.2025-2303

Ruffman, T., Henry, J. D., Livingstone, V., & Phillips, L. H. (2008). A meta-analytic review of emotion recognition and aging: Implications for neuropsychological models of aging. *Neuroscience & Biobehavioral Reviews, 32*(4), 863–881. https://doi.org/10.1016/j.neubiorev.2008.01.001

Scherer, K. R. (2003). Vocal communication of emotion: A review of research paradigms. *Speech Communication, 40*(1–2), 227–256. https://doi.org/10.1016/s0167-6393(02)00084-5

Schirmer, A., & Kotz, S. A. (2006). Beyond the right hemisphere: Brain mechanisms mediating vocal emotional processing. *Trends in Cognitive Sciences, 10*(1), 24–30. https://doi.org/10.1016/j.tics.2005.11.009

Shen, J., & Souza, P. E. (2017). The effect of dynamic pitch on speech recognition in temporally modulated noise. *Journal of Speech, Language, and Hearing Research, 60*(9), 2725–2739. https://doi.org/10.1044/2017_JSLHR-H-16-0389

Shen, J., & Souza, P. E. (2018). On dynamic pitch benefit for speech recognition in speech masker. *Frontiers in Psychology, 9,* Article 1967. https://doi.org/10.3389/fpsyg.2018.01967

Shen, J., Wright, R., & Souza, P. E. (2016). On older listeners' ability to perceive dynamic pitch. *Journal of Speech, Language, and Hearing Research, 59*(3), 572–582. https://doi.org/10.1044/2015_JSLHR-H-15-0228

Singh, G., Liskovoi, L., Launer, S., & Russo, F. (2019). The Emotional Communication in Hearing Questionnaire (EMO-CHeQ): Development and evaluation. *Ear and Hearing, 40*(2), 260–271. https://doi.org/10.1097/AUD.0000000000000611

Smeds, K., Wolters, F., & Rung, M. (2015). Estimation of signal-to-noise ratios in realistic sound scenarios. *Journal of the American Academy of Audiology, 26*(2), 183–196. https://doi.org/10.3766/jaaa.26.2.7

Song, J. H., Skoe, E., Banai, K., & Kraus, N. (2011). Perception of speech in noise: Neural correlates. *Journal of Cognitive Neuroscience, 23*(9), 2268–2279. https://doi.org/10.1162/jocn.2010.21556

Souza, P., Arehart, K., Miller, C. W., & Muralimanohar, R. K. (2011). Effects of age on *F*0 discrimination and intonation perception in simulated electric and electroacoustic hearing. *Ear and Hearing, 32*(1), 75–83. https://doi.org/10.1097/AUD.0b013e3181eccfe9

The MathWorks, Inc. (2024). *MATLAB* (Version R2023b) [Computer software]. https://www.mathworks.com

Tremblay, P., Brisson, V., & Deschamps, I. (2021). Brain aging and speech perception: Effects of background noise and talker variability. *NeuroImage, 227,* Article 117675. https://doi.org/10.1016/j.neuroimage.2020.117675

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis* (2nd ed.). Springer. https://doi.org/10.1007/978-0-387-98141-3

Wingfield, A., Tun, P. A., & McCoy, S. L. (2005). Hearing loss in older adulthood: What it is and how it interacts with cognitive performance. *Current Directions in Psychological Science, 14*(3), 144–148. https://doi.org/10.1111/j.0963-7214.2005.00356.x

Wright, A., Saxena, S., Sheppard, S. M., & Hillis, A. E. (2018). Selective impairments in components of affective prosody in neurologically impaired individuals. *Brain and Cognition, 124,* 29–36. https://doi.org/10.1016/j.bandc.2018.04.001

Wu, Y.-H., Stangl, E., Chipara, O., Hasan, S. S., Welhaven, A., & Oleson, J. (2018). Characteristics of real-world signal-to-noise ratios and speech listening situations of older adults with mild-to-moderate hearing loss. *Ear and Hearing, 39*(2), 293–304. https://doi.org/10.1097/aud.0000000000000486

Yost, W. A., & Killion, M. C. (1997). Hearing thresholds. *Encyclopedia of Acoustics, 3,* 1545–1554. https://doi.org/10.1002/9780470172537.ch123

Zhang, M., & Ding, H. (2022). Impact of background noise and contribution of visual information in emotion identification by native Mandarin speakers. *Proceedings of Interspeech, 2022,* 1993–1997. https://doi.org/10.21437/interspeech.2022-10142