J Voice. Author manuscript; available in PMC 2024 August 06.

The relationship between pitch discrimination and fundamental frequency variation: effects of singing status and vocal hyperfunction

Allison S. Aaron[a], **Defne Abur**[a,b,c], **Kalei P. Volk**[a], **J. Pieter Noordzij**[a,d], **Lauren F. Tracy**[a,d], **Cara E. Stepp**[a,d,e]

[a]Department of Speech, Language, and Hearing Sciences, Boston University, Boston, MA

[b]Department of Computational Linguistics, Centre for Language and Cognition Groningen, University of Groningen, The Netherlands

[c]Research School of Behavioral and Cognitive Neurosciences, University of Groningen, Groningen, The Netherlands

[d]Department of Otolaryngology – Head and Neck Surgery, Boston University School of Medicine, Boston, MA

[e]Department of Biomedical Engineering, Boston University, Boston, MA

Abstract

Purpose: The purpose of this study was to investigate the relationship between pitch discrimination and fundamental frequency (f_o) variation in running speech, with consideration of factors such as singing status and vocal hyperfunction (VH).

Method: Female speakers (18–69 years) with typical voices (26 non-singers; 27 singers) and speakers with VH (22 non-singers; 30 singers) completed a pitch discrimination task and read the *Rainbow Passage*. The pitch discrimination task was a two-alternative forced choice procedure, in which participants determined whether tokens were the same or different. Tokens were a pre-recorded sustained /ɑ/ of the participant's own voice and a pitch-shifted version of their sustained /ɑ/, such that the difference in f_o was adaptively modified. Pitch discrimination and *Rainbow Passage* f_o variation were calculated for each participant and compared via Pearson's correlations for each group.

Results: A significant strong correlation was found between pitch discrimination and f_o variation for non-singers with typical voices. No significant correlations were found for the other three groups, with notable restrictions in the ranges of discrimination for both singer-groups and in the range of f_o variation values for non-singers with VH.

---

§ Corresponding Author: Allison Aaron, aaaron@bu.edu.

Conclusions: Speakers with worse pitch discrimination may increase their f_0 variation to produce self-salient intonational changes, which is in contrast to previous findings from articulatory investigations. The erosion of this relationship in groups with singing training and/or with VH may be explained by the known influence of musical training on pitch discrimination or the biomechanical changes associated with VH restricting speakers' abilities to change their f_0.

Keywords

pitch perception; f_0 production; vocal hyperfunction; singing status

1 Introduction

Models of speech motor control suggest that targets of speech production are auditory and that detection of auditory errors is crucial for integration of auditory feedback into feedforward commands (Tourville & Guenther, 2011). The Directions Into Velocities of Articulators (DIVA) model of speech production posits that speakers who can better perceive fine acoustic-phonetic details will learn target regions that are spaced further apart (Guenther et al., 1998). This model is built on experimental evidence that suggests a relationship between speech perception and production. For example, individuals with better vowel discrimination demonstrate greater contrasts in their vowel productions (Fox, 1982; Franken et al., 2017; Perkell, Guenther, et al., 2004; Perkell et al., 2008). Other studies have also found evidence within the articulatory domain of speech supporting a pereption/production relationship: individuals with better perceptual abilities showed smaller auditory target regions and greater distinctions in production for various articulatory features, including sibilants (Ghosh et al., 2010; Perkell, Matthies, et al., 2004), approximants (McAllister Byun & Tiede, 2017), Dutch obstruent devoicing (Pinget et al., 2020), Illinois English /ɑ-ɔ/ (Zhang et al., 2022), and voice onset time (VOT) for stop consonants (Lindsay et al., 2022; Newman, 2003). Despite this well-established relationship within the articulatory domain, the perception/production relationship is less well-defined for voice parameters of speech production, such as pitch. Thus, several researchers have attempted to apply current models of speech motor control (based on evidence from articulation) to the voice domain (Abur et al., 2018; Castillo-Allendes, 2021; Escera et al., 2018; Lester-Smith et al., 2020; Li et al., 2021; Mollaei et al., 2019; Perkell et al., 2000). However, before models of speech motor control can be appropriately adapted for clinical voice application, we need more information about the extent to which they apply to voice, including defining the relationship between pitch discrimination and production. Specifically, further targeted research is needed to establish this relationship between auditory perceptual abilities via objective measurements of auditory acuity and acoustic characteristics of voice production.

There is indeed evidence that the control of voice production depends on the perception of auditory feedback. When auditory feedback is unavailable, mature adults demonstrate a decline in vocal control. For example, individuals with hearing impairments often exhibit diminished vocal control during habitual speech, with reduced and/or atypical fundamental frequency (f_0) variation, increased mean f_0, and/or atypical voice quality (Higgins et al., 1994; Monsen, 1983). Another well-studied and long-standing body of evidence for a relationship between voice perception and production is the involuntary increase in vocal intensity when speaking in noisy environments, known as the Lombard effect (Lombard, 1911). When masking noise attenuates auditory feedback of speech, individuals produce a robust increase in the intensity level and f_0 of the speech signal (Junqua, 1996). Finally, numerous experimental studies have used altered auditory feedback of voice f_0 (i.e., auditory feedback is experimentally manipulated such that the vocal f_0 of the feedback is shifted in real-time) to investigate the role of auditory feedback on speech production. Researchers have consistently observed compensatory responses to these perturbations of f_0 in speakers with typical voices: individuals shift their f_0 in the opposing direction of the manipulated auditory feedback (Burnett et al., 1997; Chen et al., 2007; Houde & Jordan, 1998). These findings, in both the pitch and loudness domains, suggest that individuals' auditory perceptual abilities play a crucial role in the control of voice.

A limited number of studies have specifically investigated the relationship between the perception and production of voice in speakers with typical voices. In a study conducted by Park et al. (2019), researchers examined the relationship between perception and production of breathy voice quality. They found that individuals with greater precision in categorizing typical and breathy voices had typical voices that were less breathy, as compared to individuals with lower precision. However, individuals did not show worse discrimination within-category than they did at category boundaries. This implies that perception of voice quality may not be influenced by categorical perception, in which perception of speech sounds is more precise within-category boundaries and less precise at the category boundaries (Liberman et al., 1957). Given this difference in the way speakers perceive articulation and voice quality, it is crucial that the perception/production relationship for voice is considered separately from what is known about articulatory perception and production. Furthermore, since the perception/production relationship for voice quality and pitch may not be the same, there is still a need for additional evidence that is specific to the pitch domain. One recent study investigated the relationship between pitch discrimination and acoustic measures of voice in female speakers with and without musical training (Yun et al., 2022). No significant relationships were found between pitch discrimination and various acoustic measures that included the standard deviation (SD) of f_0 during the sustained vowel /ɑ/. However, f_0 variation (f_0 SD) in running speech was not investigated. Given the prior evidence for a pereption/production relationship in articulatory motor control, it would be worthwhile to investigate pitch discrimination ability and f_0 variation within the context of running speech (i.e., f_0 SD at the sentence level as a prosodic feature of intonation, as opposed to a local-level pitch control parameter for a sustained vowel). If extant, this relationship is likely to be affected by common features of speakers that are known to influence the voice perception and production.

It is well-documented in the music cognition literature that pitch discrimination ability is influenced by musicality. That is, individuals with musical training, including singers, perform better on pitch discrimination tasks as compared to non-musicians (Kishon-Rabin et al., 2001; Micheyl et al., 2006; Nikjeh et al., 2008; Tervaniemi et al., 2005). Additionally, it has been shown that musical expertise can improve one's ability to process speech (Varnet et al., 2015) and to comprehend speech in noise (Du & Zatorre, 2017; Parbery-Clark et al., 2009). In the study by Yun et al. (2022), which compared pitch discrimination ability across musically trained and non-trained groups, the musically trained group had a significantly higher percentage of accurate responses. Further, when the musically trained group was organized by instrumentalist and vocalist groups, the vocalist group had a significantly higher percentage of accurate responses than the instrumentalist group. In a recent chapter review, existing literature on singing was applied to the DIVA model: it was suggested that singers may have a more refined auditory representation of vocal signals, particularly for their own voices, as compared to non-singers (Zuk et al., 2022). Given this known influence of musicality and singer-status on perception ability, it is important to consider musicality and singing experience when investigating the relationship between pitch discrimination and f_o variation.

Another variable that may influence the pitch perception/production relationship is whether or not the speaker has a voice disorder. Vocal hyperfunction (VH) is characterized by “excessive perilaryngeal musculoskeletal activity during phonation” (Oates & Winkworth, 2008) and is considered the most commonly diagnosed type of voice disorder (Bhattacharyya, 2014). At its core, VH is a disorder of vocal production, with reports of increased laryngeal tension resulting in changes in voice quality, fatigue, and muscular pain (Hillman et al., 2020). These changes in voice production are attributed to laryngeal biomechanics, and speakers with VH have previously been found to have reduced f_o variation during running speech as compared to speakers with typical voices (Mehta et al., 2015; Van Stan et al., 2020; Van Stan et al., 2015). Based on these findings, it would be valuable to consider the influence of this common voice condition on the production variable of interest, f_o variation in running speech, when interpreting the perception/production relationship.

In summary, a fundamental finding that has informed our knowledge of speech motor control comes from examination of articulation: individuals who have better auditory acuity to different phonemes also create greater distinctions in their phoneme productions. However, this perception/production relationship, key to current models of speech motor control, is less clearly established for voice. Further, it is unclear how this relationship is impacted by singing training (known to impact voice *perception*) and VH (known to impact voice *production*). This study aimed to investigate the relationship between pitch perception (pitch discrimination) and production (f_o variation in running speech) for individuals with VH and individuals with typical voices, with and without singing experience. We hypothesized that non-singers with typical voices would show a relationship between pitch discrimination ability and f_o variation, such that those with better discrimination would have increased f_o variation in running speech. Further, we hypothesized that this relationship would be weakened for singers and for individuals with VH secondary to an influence of singing experience on pitch discrimination ability and an influence of VH on f_o variation.

2 Methods

2.1 Participants

One hundred and five female speakers between the ages of 18–69 years (mean (M) = 29.7 years; standard deviation (SD) = 12.9 years) were included in this study. Of note, a large subset of these participants’ pitch discrimination was previously collected and published (Abur et al., 2021). Participants were organized into four groups organized by singing status (i.e., singer vs. non-singer) and presence of VH (i.e., individuals with typical voices vs. individuals with VH), with 26 non-singers with typical voices (all cisgender; M = 30.5 years, SD = 13.3 years), 27 singers with typical voices (26 cisgender, one genderqueer; M = 22.9 years, SD = 4.1 years), 22 non-singers with VH (all cisgender; M = 43.8 years, SD = 14.6 years), and 30 singers with VH (all cisgender; M = 24.6 years, SD = 7.3 years). Participants were considered singers if they had at least 5 years of formal training in vocal performance. All individuals with VH were diagnosed by a laryngologist based on a comprehensive voice evaluation, which included videolaryngoscopy at either the Boston Medical Center or the Massachusetts General Hospital Voice Center. Among the 52 individuals with VH, there were 31 with a diagnosis consistent with nonphonotraumatic VH (e.g., muscle tension dysphonia) and 21 with a diagnosis consistent with phonotraumatic VH (e.g., vocal fold nodules). Individuals with VH completed the patient-reported Voice-Related Quality of Life (V-RQOL) questionnaire (Hogikyan & Sethuraman, 1999) on the day of their experimental session (range = 10–35; M = 19.3; SD = 6.3). None of the individuals with VH had a history of neurological disorders or other speech, language, and hearing disorders. All individuals with typical voices reported no history of neurological, voice, speech, language, or hearing disorders. No participants reported use of hormone therapy or other medications that may impact the voice. Voice quality for participants with VH was rated by a blinded voice-specializing speech-language pathologist with 7 years of experience. The SLP used a visual analog scale of overall severity of dysphonia with anchors modelled after the Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V) assessment (range = 0.0–25.8; M = 10.3; SD = 10.0; Kempster et al., 2009).

Given the influence of hearing loss on auditory processing and speech perception (Lesica, 2018; Pichora-Fuller & Souza, 2003), all participants passed a hearing screening at 25 dB HL from 125 to 4000 Hz (American Speech-Language and Hearing Sciences, 2018) prior to being included in this study. All participants were fluent speakers of American English, and all participants completed written consent in compliance with the Boston University Institutional Review Board.

2.2 Procedure

Participants completed all tasks across one or two sessions each lasting 2–3 hours, which included the hearing screening, completion of voice-related surveys, and experimental tasks as part of a larger study (Abur et al., 2021), including a pitch discrimination task and a speech production reading passage task. Data from the experimental tasks were collected in a sound-attenuated booth. Participants wore an omnidirectional headset microphone (MX153; Shure, Niles, IL) placed 7 cm from the corner of their lips at a 45-degree angle (Patel et al., 2018). The microphone gain was adjusted with a preamplifier (RME Quadmic II) and the signal was digitized with a soundcard (MOTU Ultralite-mk3 Hybrid or RME Fireface UCX). Prior to the data collection session, the software and hardware systems were calibrated using a 2 cc coupler (Type 4946, Bruel and Kjaer Inc), which was connected to a sound level meter (Type 2250A with a Type 4947 ½” Pressure Field Microphone, Bruel and Kjaer). The earphone intensity output was calibrated using a 1 kHz tone played from a handheld recorder (Olympus LS-10 Linear PCM Recorder), which was positioned 7 cm from the microphone.

2.2.1 Pitch Discrimination Task—

Prior to initiation of the pitch discrimination task, participants were asked to record a sustained /ɑ/ for 3 seconds. A steady 500-ms portion from the middle of the vowel was extracted for use during the task. Participants then completed a two-alternative forced choice procedure, during which they listened to two tokens via headphones (either Etymotic ER-2 insert earphones or Sennheiser HD 280 Pro) played at a set level of 75 dB SPL and determined whether they were the “same” or “different.” The tokens included a reference stimulus (i.e., the pre-recorded 500-ms sustained /ɑ/ of the participant’s own voice that was previously extracted) and a f_o -shifted version of their sustained /ɑ/, in which the difference in f_o was adaptively modified over trials. Offline experimental shifts in voice f_o were applied to the reference stimulus to create the f_o-shifted tokens. For the majority of participants (N = 101), experimental shifts in voice f_o were applied using an Eventide Eclipse V4 Harmonizer. Due to an early technical adjustment in experimental protocol, for four participants, Audapter software (Cai et al., 2008) was used to create shifts in voice f_o. The order of the two tokens for each trial was randomized. The initial perturbation applied to the f_o-shifted token was +50 cents, with a 4-cent change in direction following two correct responses (i.e., the difference in f_o decreased by 4 cents) or one incorrect response (i.e., the difference in f_o increased by 4 cents). To ensure that participants were attending to the task, 20% of trials were “catch trials,” in which the reference stimulus was played twice. Catch trials were not included in the adaptive logic. All participants had a ≥ 63% accuracy (M = 93% accuracy, SD = 11% accuracy) on catch trials. The task was complete once the participant reached either ten reversals (i.e. changes in direction), which occurred for 92 participants, or 60 adaptive trials, which occurred for the remaining 13 participants. The average number of reversals was 9.8, and the average number of total trials was 47.2. The experiment lasted 4.3 minutes on average (SD = 0.9).

2.2.2 Acoustic Recording of the Rainbow Passage—

Participants were asked to read the first two paragraphs of the *Rainbow Passage* (Fairbanks, 1960) in their typical speaking voice. Audio recordings of the passage were recorded using Sonar Artist (Cakewalk, Boston, MA).

2.3 Data Analysis

As illustrated in Figure 1, pitch discrimination in semitones (ST) was calculated for each participant by estimating the average f_o difference value across the last six reversals (Abur & Stepp, 2020). To measure f_o variation in running speech, the mean f_o and f_o SD (both in Hz) from recordings of the *Rainbow Passage* were estimated using Praat (Boersma, 2015). The f_o settings were manually adjusted by a trained technician (K.P.V.) to optimize tracking for each participant. This trained technician reanalyzed 15% of the sample several months after the initial analysis, and intra-rater reliability of f_o SDs was calculated ($r = 0.98$) with a Pearson product-moment correlation. A second trained technician (A.S.A.) independently manually adjusted pitch settings and calculated mean f_o and f_o SD in Praat for 15% of the total dataset, and inter-rater reliability ($r = 0.99$) was calculated. Given the logarithmic relationship between f_o in Hz and pitch perception, the f_o SD was normalized to the mean f_o (ST) of each participant using Equation 1. This allows for comparisons across individuals with varied values of f_o (Hz).

$$\mathrm{ST} = 12 \times \log_2 \frac{fo\,(Hz) + fo\,SD\,(Hz)}{fo\,(\mathrm{Hz})} \qquad \text{Eq. 1}$$

2.4 Statistical Analysis

Statistical analyses were conducted in RStudio (RStudio Team, 2020). A Pearson product-moment correlation between pitch discrimination (ST) and f_0 SD (ST) was calculated for each group. Significance was set *a priori* to $p < 0.05$. Effect sizes were interpreted for statistically significant correlations, such that correlation coefficients of $r > 0.10$–29 were classified as weak, $r > 0.30$–0.49 were classified as moderate, and $r > 0.50$ were classified as strong (Cohen & Ebl, 1988).

3 Results

Summary statistics, including quartiles and medians, for both pitch discrimination and f_0 variation (f_0 SD) are presented by group in Table 1. As illustrated by the median values for pitch discrimination in Table 1, the singer groups had the best pitch discrimination (i.e., smallest estimated f_0 difference that they could perceive), followed by the non-singers with typical voices. The non-singers with VH had the worst pitch discrimination as compared to the other three groups. As for the median f_0 SD values, the singer groups had greater f_0 SD as compared to non-singers, and the non-singers with VH had the lowest f_0 SD as compared to the other three groups. A statistically significant, strong relationship between pitch discrimination and f_0 SD was observed for non-singers with typical voices ($r = .53$). Of note, this correlation coefficient was on the cusp between moderate and strong, based on the interpretation of effect size as outlined by Cohen & Ebl (1988). There were no statistically significant relationships observed for singers with typical voices, singers with VH, or non-singers with VH. Statistical results are presented in Table 2. Figure 2 shows scatterplots of pitch discrimination and f_0 SD for all four groups.

4 Discussion

We hypothesized that non-singers with typical voices would show a relationship between pitch discrimination ability and f_0 variation, such that those with better discrimination would have increased f_0 variation in running speech. This hypothesis was based on findings that support a perception and production relationship in the articulatory domain, in which better articulatory discrimination was associated with increased produced articulatory contrasts (Ghosh et al., 2010; Lindsay et al., 2022; McAllister Byun & Tiede, 2017; Newman, 2003; Perkell, Guenther, et al., 2004; Perkell, Matthies, et al., 2004; Pinget et al., 2020; Zhang et al., 2022). Instead, the current study found a significant strong correlation between pitch discrimination and f_0 variation for non-singers with typical voices, in which worse pitch discrimination was associated with increased produced f_0 variation. This finding is opposite our *a priori* hypothesis, that better pitch discrimination would be associated with increased produced f_0 variation. One interpretation of this relationship is that individuals with worse discrimination have greater difficulty discriminating between f_0 changes in their own voice. Therefore, to produce self-salient intonational changes, they must increase their f_0 variation.

The DIVA model is based on experimental work that supports the notion that speech production depends upon auditory perception, and that targets for speech and voice are both auditory and categorical (Guenther et al., 1998; Tourville & Guenther, 2011). There have been several attempts to apply the DIVA model and similar models of speech motor control for the benefit of populations with voice disorders. This includes a publication of recommendations for emphasizing the processes of auditory-vocal integration in assessment and treatment of voice disorders (Castillo-Allendes, 2021). Therapy programs, such as Lee Silverman Voice Treatment (LSVT® Loud), have been suggested to improve auditory-vocal integration of vocal pitch production in individuals with Parkinson's Disease (Li et al., 2021), and research on commercially available clinical devices that alter auditory feedback, such as Forbrain® (Escera et al., 2018) with the intention of enhancing auditory-motor processing and integration. However, based on the preliminary finding from this study, there is reason to consider future investigations of pitch perception/production relationships separately from interpretations surrounding the articulatory domain. Unlike articulation, which has been shown to be influenced by categorical perception (Kuhl, 2004; Liberman et al., 1957), perception of pitch may be continuous within a musical context for non-musicians (Burns & Ward, 1978; Zarate & Zatorre, 2005). This is contrasted by the categorical perception of pitch observed for trained musicians in the Western music tradition (Burns & Ward, 1978; Siegel & Siegel, 1977; Sundberg, 1994; Zarate et al., 2012). Continuous perception of pitch has also been shown within a context of tonal language categories for non-native speakers, as compared to native-speakers of tonal languages (Francis et al., 2008; Peng et al., 2010; Shen & Froud, 2016; Xu et al., 2006). Given this evidence that suggests that perceptual mechanisms for articulation and pitch are inherently different from one another, to better model f_o control for speech production, more research is needed to clarify the nuances of pitch perception and factors that may influence it. One prior study provided evidence for a pereption/production relationship for breathy voice quality, in which individuals with better perceptual precision in differentiating between typical and breathy voices produced their own voices with less breathiness (Park et al., 2019). However, despite this relationship, individuals did not show worse discrimination within-category than they did at category boundaries, a finding that is integral to the distinction between categorical and continuous perception and observed in the articulatory literature (Kuhl, 2004; Liberman et al., 1957). This finding suggests that voice quality may not be perceived categorically, but is instead continuously. It is possible that perceptual mechanisms within the voice domain, including both voice quality and pitch are comparable to one another. Because the mechanisms of perception of articulation may differ significantly from both voice quality and pitch perception, the present study finding offers preliminary evidence that is required for the appropriate and independent modeling of f_0 control in speech production. This study provides novel information regarding pereption/production relationships within the voice domain (i.e., pitch) that is crucial for our overall understanding and successful application of concepts to enhance motor learning in the voice clinics. Future research is needed so that we can appropriately identify specific targets for voice application.

We hypothesized that the pitch pereption/production relationship would be weakened for singers due to an influence of singing experience on pitch discrimination ability, and weakened for individuals with VH, due to an influence of VH on f_0 variation. As expected, there were no significant correlations for singers with typical voices, singers with VH, and non-singers with VH. As shown in the distributions in Table 1, there was a notable restriction in the range of pitch discrimination for singers with typical voices (range: 0.10–0.54 ST) and singers with VH (range: 0.03–0.60 ST), as compared to non-singers with typical voices (range: 0.14–0.88 ST). Also, there was a notable restriction in the range of f_0 SD values for non-singers with VH (range: 1.03–2.74 ST) as compared to non-singers with typical voices (range: 1.12–3.56 ST).

It is possible that the correlation found in the non-singers with typical voices was eroded by the known influence of musicality on pitch discrimination in the singers with typical voices and singers with VH groups, demonstrated by their restricted range of pitch discrimination. The lack of relationships for both singer groups supports the second study hypothesis, and aligns with prior research that shows that trained musicians perform better on pitch discrimination tasks, as compared to non-musicians (Micheyl et al., 2006). Of note, there are also normative data to suggest that the f_0 production variable may also be influenced by singing status. That is, singers may have a higher f_0 and greater f_0 SD in running speech as compared to non-singers (Baken, 1987; Colton et al., 2011; Siupsinskiene & Lycke, 2011), which contributes to the complexity of interpreting a pitch perception/production relationship in a group of singers. Additionally, The Linked Dual Representation model of vocal perception and production (Hutchins & Moreno, 2013) interprets conflicting literature on the relationship between vocal perception and production in the context of singing abilities. The authors state that prior research points to evidence of a link between perception and production for singing. However, there is additional growing evidence for a dissociation between vocal perception and production, including for individuals with tone deafness (i.e. congenital amusia) and those with extensive singing training. When interpreting the findings of the current study, this model may help to interpret why there was a correlation between perception and production for non-singers, but not for the other groups in this study. This model strengthens the importance of considering factors, such as singing-status, that may increase instances in which an individuals' production abilities outstrip their perception abilities, or vice versa.

In addition to the influence of singing status on the relationship between pitch discrimination and f_0 variation, we also hypothesized that the relationship would be weakened for individuals with VH. This hypothesis was based on the voice changes associated with VH that would restrict speakers' abilities to vary their f_o. This was consistent with the results from this study, with no relationship found for non-singers with VH, and a restricted range of f_0 SD values. This interpretation aligns with prior work that considers altered laryngeal biomechanics during phonation as a precipitating factor for development of VH (Hillman et al., 2020). It is also supported by ambulatory voice monitoring data that shows decreased f_0 variation as an acoustic feature that distinguishes a group of individuals with VH from individuals with typical voices (Van Stan et al., 2020). Of note, according to the framework developed by Hillman and colleagues' (2020), the etiology of VH is heterogeneous, and includes factors attributed to personality, sensorimotor deficits and anatomical/physiological vulnerability. In fact, there is even preliminary evidence for increased prevalence of undiagnosed hearing impairment in individuals with VH (Nagy et al., 2020), and a recent study by Abur et al. (2021) found poorer pitch discrimination tasks in individuals with VH as compared to controls with typical voices, which includes a subsection of the data from the current study. Given this added complexity, it may be important in future studies to consider not only the impact of VH on f_0 production, but also on perception.

This research investigated sentence-level f_0 production for connected speech, rather than a sustained vowel production, to capture individual f_0 variation. This is crucial, given that f_0 variation of a sustained vowel production is not included in gold-standard acoustic evaluation protocols (Patel et al., 2018), and it does not allow us to generalize findings to communicative speech production. Instead, sentence-level f_0 variation allows for valuable ecologically valid intonational information to be measured from the speech sample. However, a limitation of this study and its connections to models of speech motor control is the comparison of a local variable of pitch discrimination at the phoneme-level with a global variable of f_0 variation at the sentence-level. Additionally, speech production was measured for a standard reading passage, rather than for spontaneous speech. This method allowed for consistent, structured comparison across participants and groups without introducing phonemic, linguistic, or prosodic variability into the stimuli, and is supported by published acoustic evaluation recommendations (Patel et al., 2018). However, the $f0$ variation values should be interpreted and applied to conversational speech with caution, given that mean f_0 and/or f_0 variation during speech production of a reading passage may vary from measures of f_0 for spontaneous speech. (Graddol, 2018; Zraick et al., 2000). Another limitation of this study is that the singer groups were, on average, younger than the non-singer groups. Future research should control for age-related changes to perception and production of pitch, by comparing more closely age-matched singer and non-singer groups.

In this study, we did not collect data on the specifics of singers' musical training. Given the known musical differences in the thresholds of pitch intervals between Western and Eastern trained musicians (Zarate et al., 2012), future work could investigate the relationship between pitch discrimination and f_0 variation in musicians who may be trained outside of the Western musical tradition, as these individuals are often trained to perceive and produce pitch intervals smaller than the typical Western musical threshold of one semitone. Additionally, future work should include a perceptual task that measures perception at the sentence-level, to better match the f_0 production variable, which is designed as a measure of intonation. Related studies have investigated the normative f_0 difference required to perceive linguistic emphasis (Rietveld & Gussenhovent, 1985) and explored the mechanisms surrounding perception of intonation for English (Dilley, 2010; Ladd & Morton, 1997; Roy et al., 2017). However, further research is needed to determine whether there are individual and/or group differences for perception of intonation, and a corresponding relationship with production.

5 Conclusions

The present study investigated the relationship between pitch perception and production in individuals with typical voices, and in individuals with singing experience and/or a diagnosis of VH. Results for non-singers with typical voices indicate that those with worse pitch discrimination abilities produce greater intonational changes, as measured by f_0 variation, during paragraph-level connected speech. This is not consistent with previous articulatory work, which has shown that individuals with better articulatory discrimination abilities produce greater articulatory contrasts (Fox, 1982; Franken et al., 2017; Ghosh et al., 2010; Lindsay et al., 2022; McAllister Byun & Tiede, 2017; Newman, 2003; Perkell, Guenther, et al., 2004; Perkell et al., 2008; Perkell, Matthies, et al., 2004; Pinget et al., 2020; Zhang et al., 2022). There were no significant correlations between pitch discrimination and f_0 variation for singers or for individuals with VH. Thus, this study provides preliminary evidence for a pereption/production relationship for voice in non-singers with typical voices and demonstrates that variables such as singing status and voice disorder may impact that relationship. These findings suggest that future research is needed before models of speech motor control based on articulation are directly applied to the voice domain. Future research should further investigate the relationship between perception and production for f_0 control, with specific attention to individual and/or group differences for perception and production at the sentence-level.

Acknowledgments

This work was supported by the National Institute on Deafness and Other Communication Disorders grants DC015446 (R. E. H.), DC013017 (C. A. M. and C. E. S.), and F31 DC019032 (D.A.). This work was also supported by an ASHFoundation New Century Doctoral Scholarship (D.A.) and a Graduate Fellow Award from the Rafik B. Hariri Institute for Computing and Computational Science and Engineering (D.A.). The authors thank Alyssa Williams for assistance with data processing.

7 References

Abur D, Lester-Smith RA, Daliri A, Lupiani AA, Guenther FH, & Stepp CE (2018). Sensorimotor adaptation of voice fundamental frequency in Parkinson's disease. Plos One, 13(1), e0191839. [PubMed: 29373589]

Abur D, & Stepp CE (2020). Acuity to Changes in Self-Generated Vocal Pitch in Parkinson's Disease. J Speech Lang Hear Res, 63(9), 3208–3214. 10.1044/2020_JSLHR-20-00003 [PubMed: 32853119]

Abur D, Subaciute A, Kapsner-Smith M, Segina RK, Tracy LF, Noordzij JP, & Stepp CE (2021). Impaired auditory discrimination and auditory-motor integration in hyperfunctional voice disorders. Sci Rep, 11(1), 13123. 10.1038/s41598-021-92250-8 [PubMed: 34162907]

American Speech-Language and Hearing Sciences. (2018). Scope of practice in audiology [Scope of practice]. www.asha.org/policy/.

Baken RJ (1987). Clinical measurement of speech and voice. Taylor & Francis.

Bhattacharyya N (2014). The prevalence of voice problems among adults in the United States. Laryngoscope, 124(10), 2359–2362. 10.1002/lary.24740 [PubMed: 24782443]

Boersma PW,D (2015). Doing phonetics by computer. http://www.praat.org

Burnett TA, Senner JE, & Larson CR (1997). Voice F0 responses to pitch-shifted auditory feedback: a preliminary study. Journal of Voice, 11(2), 202–211. 10.1016/s0892-1997(97)80079-3 [PubMed: 9181544]

Burns EM, & Ward WD (1978). Categorical perception—phenomenon or epiphenomenon: Evidence from experiments in the perception of melodic musical intervals. The Journal of the Acoustical Society of America, 63(2), 456–468. [PubMed: 670543]

Cai S, Boucek M, Ghosh SS, Guenther FH, & Perkell JS (2008). A system for online dynamic perturbation of formant trajectories and results from perturbations of the Mandarin triphthong/iau. Proceedings of the 8th ISSP, 65–68.

Castillo-Allendes A, Contreras-Ruston F, & Searl J (2021). Auditory-vocal integration impairment: New challenges and opportunities for voice assessment and therapy. Revista De Investigación E Innovación En Ciencias De La Salud, 3(2), 87–97. https://doi.org/10.46634/riics.62

Chen SH, Liu H, Xu Y, & Larson CR (2007). Voice F0 responses to pitch-shifted voice feedback during English speech. Journal of the Acoustical Society of America, 121(2), 1157–1163. 10.1121/1.2404624 [PubMed: 17348536]

Cohen J, & Ebl. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). L. Erlbaum Associates.

Colton RH, Casper JK, & Leonard R (2011). Understanding voice problems : a physiological perspective for diagnosis and treatment (Fourth edition. ed.). LWW. http://whel-primo.hosted.exlibrisgroup.com/openurl/44WHELF_NLW/44WHELF_NLW_services_page?u.ignore_date_coverage=true&rft.mms_id=99953828302419 https://tcdlibrary.ldls.org.uk/vdc_100031111126.0x000001

Dilley LC (2010). Pitch range variation in English tonal contrasts: Continuous or categorical? Phonetica, 67(1-2), 63–81. [PubMed: 20798570]

Du Y, & Zatorre RJ (2017). Musical training sharpens and bonds ears and tongue to hear speech better. Proc Natl Acad Sci U S A, 114(51), 13579–13584. 10.1073/pnas.1712223114 [PubMed: 29203648]

Escera C, Lopez-Caballero F, & Gorina-Careta N (2018). The Potential Effect of Forbrain as an Altered Auditory Feedback Device. J Speech Lang Hear Res, 61(4), 801–810. 10.1044/2017_JSLHR-S-17-0072 [PubMed: 29554188]

Fairbanks G (1960). Voice and articulation drillbook (2nd ed.). Harper;H.Hamilton.

Fox RA (1982). Individual variation in the perception of vowels: Implications for a pereption/production link. Phonetica, 39(1), 1–22. [PubMed: 7089067]

Francis AL, Ciocca V, Ma L, & Fenn K (2008). Perceptual learning of Cantonese lexical tones by tone and non-tone language speakers. Journal of phonetics, 36(2), 268–294.

Franken MK, Acheson DJ, McQueen JM, Eisner F, & Hagoort P (2017). Individual variability as a window on production-perception interactions in speech motor control. The Journal of the Acoustical Society of America, 142(4), 2007–2018. [PubMed: 29092613]

Ghosh SS, Matthies ML, Maas E, Hanson A, Tiede M, Menard L, Guenther FH, Lane H, & Perkell JS (2010). An investigation of the relation between sibilant production and somatosensory and auditory acuity. Journal of the Acoustical Society of America, 128(5), 3079–3087. 10.1121/1.3493430 [PubMed: 21110603]

Graddol D (2018). Discourse specific pitch behaviour. In Intonation in discourse (pp. 221–238). Routledge.

Guenther FH, Hampson M, & Johnson D (1998). A theoretical investigation of reference frames for the planning of speech movements. Psychol Rev, 105(4), 611–633. 10.1037/0033-295x.105.4.611-633 [PubMed: 9830375]

Higgins MB, Carney AE, & Schulte L (1994). Physiological Assessment of Speech and Voice Production of Adults With Hearing Loss. Journal of Speech, Language, and Hearing Research, 37(3), 510–521. 10.1044/jshr.3703.510

Hillman RE, Stepp CE, Van Stan JH, Zanartu M, & Mehta DD (2020). An Updated Theoretical Framework for Vocal Hyperfunction. Am J Speech Lang Pathol, 29(4), 2254–2260. 10.1044/2020_AJSLP-20-00104 [PubMed: 33007164]

Hogikyan ND, & Sethuraman G (1999). Validation of an instrument to measure voice-related quality of life (V-RQOL). Journal of Voice, 13(4), 557–569. 10.1016/s0892-1997(99)80010-1 [PubMed: 10622521]

Houde JF, & Jordan MI (1998). Sensorimotor adaptation in speech production. Science, 279(5354), 1213–1216. 10.1126/science.279.5354.1213 [PubMed: 9469813]

Hutchins S, & Moreno S (2013). The Linked Dual Representation model of vocal perception and production. Frontiers in Psychology, 4, 825. [PubMed: 24204360]

Junqua JC (1996). The influence of acoustics on speech production: A noise-induced stress phenomenon known as the Lombard reflex. Speech Communication, 20(1-2), 13–22. https://doi.org/Doi10.1016/S0167-6393(96)00041-6

Kempster GB, Gerratt BR, Verdolini Abbott K, Barkmeier-Kraemer J, & Hillman RE (2009). Consensus auditory-perceptual evaluation of voice: development of a standardized clinical protocol. Am J Speech Lang Pathol, 18(2), 124–132. 10.1044/1058-0360(2008/08-0017) [PubMed: 18930908]

Kishon-Rabin L, Amir O, Vexler Y, & Zaltz Y (2001). Pitch discrimination: are professional musicians better than non-musicians? Journal of basic and clinical physiology and pharmacology, 12(2), 125–144. [PubMed: 11605682]

Kuhl PK (2004). Early language acquisition: cracking the speech code. Nature reviews neuroscience, 5(11), 831–843. [PubMed: 15496861]

Ladd DR, & Morton R (1997). The perception of intonational emphasis: continuous or categorical? Journal of phonetics, 25(3), 313–342.

Lesica NA (2018). Why do hearing aids fail to restore normal auditory perception? Trends in neurosciences, 41(4), 174–185. [PubMed: 29449017]

Lester-Smith RA, Daliri A, Enos N, Abur D, Lupiani AA, Letcher S, & Stepp CE (2020). The relation of articulatory and vocal auditory–motor control in typical speakers. Journal of Speech, Language, and Hearing Research, 63(11), 3628–3642.

Li Y, Tan M, Fan H, Wang EQ, Chen L, Li J, Chen X, & Liu H (2021). Neurobehavioral Effects of LSVT((R)) LOUD on Auditory-Vocal Integration in Parkinson's Disease: A Preliminary Study. Front Neurosci, 15, 624801. 10.3389/fnins.2021.624801 [PubMed: 33716652]

Liberman AM, Harris KS, Hoffman HS, & Griffith BC (1957). The discrimination of speech sounds within and across phoneme boundaries. Journal of experimental psychology, 54(5), 358. [PubMed: 13481283]

Lindsay S, Clayards M, Gennari S, & Gaskell MG (2022). Plasticity of categories in speech perception and production. Language, Cognition and Neuroscience, 1–25.

Lombard E (1911). Le signe de l’ élévation de la voix [The sign of voice raising]. Annales des Maladies de l’ Oreille et du Larynx(37), 101–119.

McAllister Byun T, & Tiede M (2017). Pereption/production relations in later development of American English rhotics. Plos One, 12(2), e0172022. 10.1371/journal.pone.0172022 [PubMed: 28207800]

Mehta DD, Van Stan JH, Zañartu M, Ghassemi M, Guttag JV, Espinoza VM, Cortés JP, Cheyne HA, & Hillman RE (2015). Using ambulatory voice monitoring to investigate common voice disorders: Research update. Frontiers in bioengineering and biotechnology, 3, 155. [PubMed: 26528472]

Micheyl C, Delhommeau K, Perrot X, & Oxenham AJ (2006). Influence of musical and psychoacoustical training on pitch discrimination. Hear Res, 219(1-2), 36–47. 10.1016/j.heares.2006.05.004 [PubMed: 16839723]

Mollaei F, Shiller DM, Baum SR, & Gracco VL (2019). The relationship between speech perceptual discrimination and speech production in Parkinson's disease. Journal of Speech, Language, and Hearing Research, 62(12), 4256–4268.

Monsen RB (1983). Voice quality and speech intelligibility among deaf children. Am Ann Deaf, 128(1), 12–19. 10.1353/aad.2112.0015 [PubMed: 6837383]

Nagy A, Elshafei R, & Mahmoud S (2020). Correlating Undiagnosed Hearing Impairment with Hyperfunctional Dysphonia. Journal of Voice, 34(4), 616–621. 10.1016/j.jvoice.2019.02.002 [PubMed: 30792081]

Newman RS (2003). Using links between speech perception and speech production to evaluate different acoustic metrics: a preliminary report. Journal of the Acoustical Society of America, 113(5), 2850–2860. 10.1121/1.1567280 [PubMed: 12765401]

Nikjeh DA, Lister JJ, & Frisch SA (2008). Hearing of note: an electrophysiologic and psychoacoustic comparison of pitch discrimination between vocal and instrumental musicians. Psychophysiology, 45(6), 994–1007. [PubMed: 18778322]

Oates J, & Winkworth A (2008). Current knowledge, controversies and future directions in hyperfunctional voice disorders. Int J Speech Lang Pathol, 10(4), 267–277. 10.1080/17549500802140153 [PubMed: 20840042]

Parbery-Clark A, Skoe E, Lam C, & Kraus N (2009). Musician enhancement for speech-in-noise. Ear Hear, 30(6), 653–661. 10.1097/AUD.0b013e3181b412e9 [PubMed: 19734788]

Park Y, Perkell JS, Matthies ML, & Stepp CE (2019). Categorization in the Perception of Breathy Voice Quality and Its Relation to Voice Production in Healthy Speakers. J Speech Lang Hear Res, 62(10), 3655–3666. 10.1044/2019_JSLHR-S-19-0048 [PubMed: 31525305]

Patel RR, Awan SN, Barkmeier-Kraemer J, Courey M, Deliyski D, Eadie T, Paul D, Svec JG, & Hillman R (2018). Recommended Protocols for Instrumental Assessment of Voice: American Speech-Language-Hearing Association Expert Panel to Develop a Protocol for Instrumental Assessment of Vocal Function. Am J Speech Lang Pathol, 27(3), 887–905. 10.1044/2018_AJSLP-17-0009 [PubMed: 29955816]

Peng G, Zheng H-Y, Gong T, Yang R-X, Kong J-P, & Wang WS-Y (2010). The influence of language experience on categorical perception of pitch contours. Journal of phonetics, 38(4), 616–624.

Perkell JS, Guenther FH, Lane H, Matthies ML, Perrier P, Vick J, Wilhelms-Tricarico R, & Zandipour M (2000). A theory of speech motor control and supporting data from speakers with normal hearing and with profound hearing loss. Journal of phonetics, 28(3), 233–272.

Perkell JS, Guenther FH, Lane H, Matthies ML, Stockmann E, Tiede M, & Zandipour M (2004). The distinctness of speakers' productions of vowel contrasts is related to their discrimination of the contrasts. Journal of the Acoustical Society of America, 116(4 Pt 1), 2338–2344. 10.1121/1.1787524 [PubMed: 15532664]

Perkell JS, Lane H, Ghosh S, Matthies ML, Tiede M, Guenther F, & Ménard L (2008). Mechanisms of vowel production: auditory goals and speaker acuity. Proceedings of the Eighth international Seminar on speech production, Strasbourg, France,

Perkell JS, Matthies ML, Tiede M, Lane H, Zandipour M, Marrone N, Stockmann E, & Guenther FH (2004). The distinctness of speakers' /s/-/S/ contrast is related to their auditory discrimination and use of an articulatory saturation effect. J Speech Lang Hear Res, 47(6), 1259–1269. 10.1044/1092-4388(2004/095) [PubMed: 15842009]

Pichora-Fuller MK, & Souza PE (2003). Effects of aging on auditory processing of speech. Int J Audiol, 42 Suppl 2, 2S11–16. https://www.ncbi.nlm.nih.gov/pubmed/12918623 [PubMed: 12918623]

Pinget AF, Kager R, & Van de Velde H (2020). Linking Variation in Perception and Production in Sound Change: Evidence from Dutch Obstruent Devoicing. Lang Speech, 63(3), 660–685. 10.1177/0023830919880206 [PubMed: 31623510]

Rietveld AC, & Gussenhovent C (1985). On the relation between pitch excursion size and prominence. Journal of phonetics, 13(3), 299–308.

Roy J, Cole J, & Mahrt T (2017). Individual differences and patterns of convergence in prosody perception. Laboratory Phonology, 8(1).

RStudio Team. (2020). RStudio: Integrated Development for R. In RStudio, PBC, Boston, MA. http://www.rstudio.com/

Shen G, & Froud K (2016). Categorical perception of lexical tones by English learners of Mandarin Chinese. The Journal of the Acoustical Society of America, 140(6), 4396–4403. [PubMed: 28040029]

Siegel JA, & Siegel W (1977). Categorical perception of tonal intervais: musicians can't tellsharp fromflat. Perception & Psychophysics, 21(5), 399–407.

Siupsinskiene N, & Lycke H (2011). Effects of vocal training on singing and speaking voice characteristics in vocally healthy adults and children based on choral and nonchoral data. Journal of Voice, 25(4), e177–189. 10.1016/j.jvoice.2010.03.010 [PubMed: 20702062]

Sundberg J (1994). Perceptual aspects of singing. Journal of Voice, 8(2), 106–122. [PubMed: 8061767]

Tervaniemi M, Just V, Koelsch S, Widmann A, & Schröger E (2005). Pitch discrimination accuracy in musicians vs non-musicians: an event-related potential and behavioral study. Experimental Brain Research, 161(1), 1–10. [PubMed: 15551089]

Tourville JA, & Guenther FH (2011). The DIVA model: A neural theory of speech acquisition and production. Lang Cogn Process, 26(7), 952–981. 10.1080/01690960903498424 [PubMed: 23667281]

Van Stan JH, Mehta DD, Ortiz AJ, Burns JA, Toles LE, Marks KL, Vangel M, Hron T, Zeitels S, & Hillman RE (2020). Differences in weeklong ambulatory vocal behavior between female patients with phonotraumatic lesions and matched controls. Journal of Speech, Language, and Hearing Research, 63(2), 372–384.

Van Stan JH, Mehta DD, Zeitels SM, Burns JA, Barbu AM, & Hillman RE (2015). Average ambulatory measures of sound pressure level, fundamental frequency, and vocal dose do not differ between adult females with phonotraumatic lesions and matched control subjects. Annals of Otology, Rhinology & Laryngology, 124(11), 864–874. [PubMed: 26024911]

Varnet L, Wang T, Peter C, Meunier F, & Hoen M (2015). How musical expertise shapes speech perception: evidence from auditory classification images. Sci Rep, 5, 14489. 10.1038/srep14489 [PubMed: 26399909]

Xu Y, Gandour JT, & Francis AL (2006). Effects of language experience and stimulus complexity on the categorical perception of pitch direction. The Journal of the Acoustical Society of America, 120(2), 1063–1074. [PubMed: 16938992]

Yun EW, Nguyen DD, Carding P, Hodges NJ, Chacon AM, & Madill C (2022). The Relationship Between Pitch Discrimination and Acoustic Voice Measures in a Cohort of Female Speakers. Journal of Voice. 10.1016/j.jvoice.2022.02.015

Zarate JM, Ritson CR, & Poeppel D (2012). Pitch-interval discrimination and musical expertise: Is the semitone a perceptual boundary? The Journal of the Acoustical Society of America, 132(2), 984–993. [PubMed: 22894219]

Zarate JM, & Zatorre RJ (2005). Neural substrates governing audiovocal integration for vocal pitch regulation in singing. Ann N Y Acad Sci, 1060, 404–408. 10.1196/annals.1360.058 [PubMed: 16597793]

Zhang JNF, Graham L, Barlaz M, & Hualde JI (2022). Within-Speaker Perception and Production of Two Marginal Contrasts in Illinois English. Frontiers in Communication, 7. 10.3389/fcomm.2022.844862

Zraick RI, Skaggs SD, & Montague JC (2000). The effect of task on determination of habitual pitch. Journal of Voice, 14(4), 484–489. [PubMed: 11130106]

Zuk J, Loui P, & Guenther F (2022). Neural Control of Speaking and Singing: The DIVA Model for Singing. In. https://doi.org/10.31234/osf.io/xqtc9

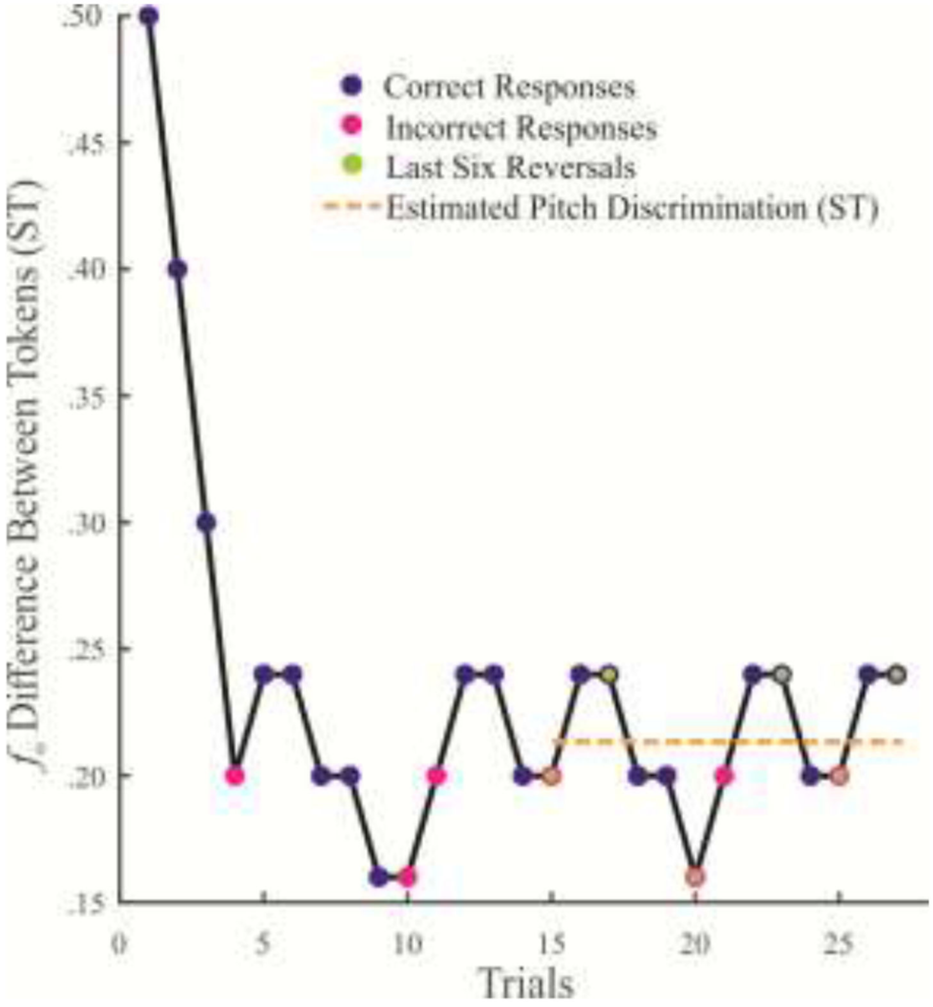

Figure 1.

Example experimental run of the pitch discrimination task for one participant. Data points represent the f_0 difference between tokens (ST) that was adaptively modified as trials progressed. The f_0 difference decreased following two correct responses or increased following one incorrect response. The dotted orange line indicates the estimated pitch discrimination in semitones (ST), as estimated by the average f_0 difference between tokens across the last six reversals.

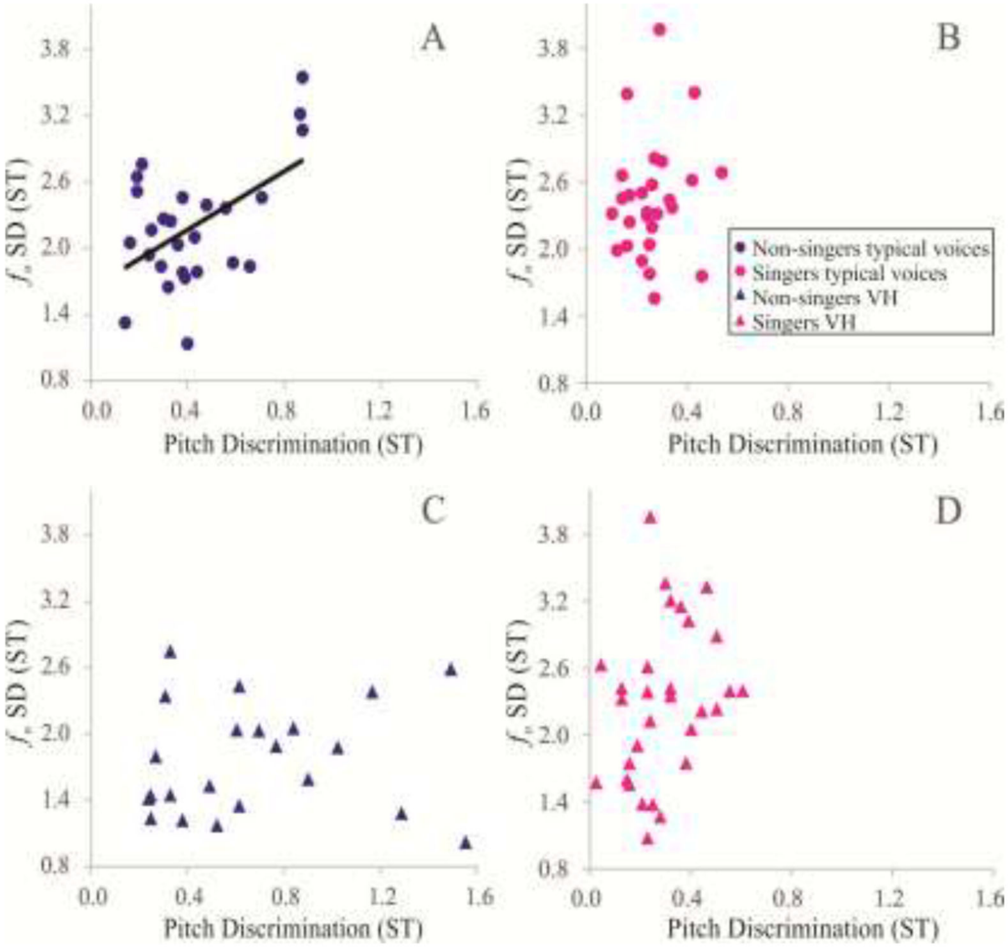

Figure 2.

Scatter plots of participant values of fundamental frequency (f_o) standard deviation (SD) in semitones (ST) and pitch discrimination in ST for (A) Non-singers with typical voices, (B) Singers with typical voices, (C) Non-singers with vocal hyperfunction (VH), and (D) Singers with VH. A line of best fit is shown for the statistically significant correlation. Data from non-singers are presented in circles and singers are presented in triangles. Data from individuals with typical voices are presented in dark blue, and individuals with VH are presented in light pink.

Table 1.

Summary statistics, including lower quartile (Q1), median, and upper quartile (Q3), for pitch discrimination in semitones (ST) and fundamental frequency (f_o) variation (standard deviation [SD] in ST) for all 105 participants separated by group based on singing- and voice disorder-status.

Group	***N***	***Pitch Discrimination* (ST)**			**f_o *Variation* (f_o SD in ST)**		
		Q1	***Median***	***Q3***	***Q1***	***Median***	***Q3***
Non-singers with typical voices	26	0.26	0.38	0.54	1.83	2.13	2.46
Singers with typical voices	27	0.17	0.25	0.30	2.11	2.37	2.64
Non-singers with vocal hyperfunction	22	0.34	0.62	0.89	1.38	1.7	2.05
Singers with vocal hyperfunction	30	0.20	0.27	0.39	1.76	2.34	2.62

Table 2.

Pearson correlations (r) for pitch discrimination in semitones (ST) and fundamental frequency standard deviation (ST) for all four groups. *Statistically significant correlation at $p < .05$.

Group	*N*	*r*	*p*
Non-singers with typical voices	26	**0.53**	**.006***
Singers with typical voices	27	0.15	.458
Non-singers with vocal hyperfunction	22	0.09	.675
Singers with vocal hyperfunction	30	0.31	.095