The Relationship Between Pitch Discrimination and Acoustic Voice Measures in a Cohort of Female Speakers

*,†Emily Wing-Tung Yun, *,†Duy Duong Nguyen, ‡Paul Carding, §Nicola J. Hodges, *,†Antonia Margarita Chacon, and *,†Catherine Madill, *†*Sydney, Australia,* ‡*Oxford, England, and* §*Vancouver, Canada*

Summary: Background. Evidence across a range of musically trained, hearing disordered and voice disordered populations present conflicting results regarding the relationship between pitch discrimination (PD) and voice quality. PD characteristics of female speakers with and without a musical training background and no self-reported voice disorder, and the relationship between PD and voice quality in this particular population, have not been investigated.

Aims. To evaluate PD characteristics in a cohort of female participants without a self-reported voice disorder and the relationship between PD and acoustic voice measures.

Method. One hundred fourteen female participants were studied, all of whom self-reported as being non-voice disordered. All completed the Newcastle Assessment of Pitch Discrimination which involved a two-tone PD task. Their voices were recorded producing standardized vocal tasks. Voice samples were acoustically analyzed for frequency-domain measures (fundamental frequency and its standard deviation, and harmonics-to-noise ratio) and spectral-domain measures (cepstral peak prominence and the Cepstral/Spectral Index of Dysphonia). Data were analyzed for the whole cohort and for musical and non-musical training backgrounds.

Results. In the whole cohort, there were no significant correlations between PD and acoustic voice measures. PD accuracy in musically trained speakers was better than in non-trained speakers and correlated with fundamental frequency standard deviation in prolonged vowel tasks. Vocalists demonstrated superior PD accuracy and fundamental frequency standard deviation in prolonged vowels compared to instrumentalists but did not show significant correlations between PD and acoustic measures. The Newcastle Assessment of Pitch Discrimination was a reliable tool, showing moderate-good prediction value in differentiating musical background.

Conclusions. There was little evidence of a relationship between PD and acoustic measures of voice quality, regardless of musical training background and superior PD accuracy among the musically trained. These data do not support ideas concerning the co-development of perception and action among individuals identified as having voice quality measures within normal ranges. Numerous measures of voice quality, including measures sensitive to pitch, did not distinguish across musically and non-musically trained individuals, despite individual differences in pitch discrimination.

Key Words: Pitch discrimination—Auditory perception—Voice quality—Musical training.

INTRODUCTION

The ability to control, change or improve vocal production depends upon voice perception,[1] which involves judgement of vocal pitch, intensity, and quality.[2] This dependency of voice production upon auditory perception has been shown in the use of auditory feedback for learning and self-monitoring in singing and vocal training[3–5] and the adjustment response seen during vocalisation in pitch perturbation studies.[6–9] Relationships between perception and production can be the result of bidirectional influences, including the effects of perception on production and the effects of production on perception. Pitch discrimination (PD), that is, the ability to tell the difference between pitches, has been frequently used to test auditory perception function.[10,11]

The role of auditory perception in voice production

The influence of perception on voice/speech production can be demonstrated typically in the hearing impaired. Habitual vocal production of individuals with hearing impairments is often associated with reduced pitch control, such as a lack of pitch variation or excessive abnormal pitch variations.[12] Moderate-profound hearing loss has also been associated with abnormal perceptual voice quality such as breathiness, hoarseness, and higher speaking fundamental frequency (F0).[13] Whilst this unidirectional link between auditory abilities and vocal production abilities has been observed, it is unknown whether a deficit in central auditory processing or a degraded peripheral auditory signal were the key factors causing reduced voice quality and control.

A relationship is also demonstrated in evidence that musicians have demonstrated better pitch production accuracy than non-musicians.[14] Trained singers also performed better

Accepted for publication February 13, 2022.

From the *Discipline of Speech Pathology, Faculty of Medicine and Health, Sydney School of Health Sciences, The University of Sydney, Sydney, Australia; †Doctor Liang Voice Program, Faculty of Medicine and Health, Sydney School of Health Sciences, The University of Sydney, Sydney, Australia; ‡Oxford Institute of Nursing, Midwifery and Allied Health Research, Oxford Brookes University, Oxford, England; and the §School of Kinesiology, University of British Columbia, Vancouver, British Columbia, Canada.

Address correspondence and reprint requests to Catherine Madill, Faculty of Medicine and Health, Sydney School of Health Sciences, The University of Sydney, Sydney, Australia. E-mail: cate.madill@sydney.edu.au

Journal of Voice, Vol. ■■, No. ■■, pp. ■■−■■

0892-1997

https://doi.org/10.1016/j.jvoice.2022.02.015

in pitch production than untrained individuals.[15] However, pitch perception and pitch production abilities were not significantly different between vocalists and musicians.[14,16] Positive correlations have also been observed between PD and vocal production in musicians.[14] Smaller difference limens for frequency (DLFs), indicative of better PD, were correlated with better pitch matching vocal accuracy in highly trained, working musicians.[14] However, correlations were not observed in non-musicians. In addition, in inaccurate singers, no relationship has been observed between PD and pitch production ability.[17]

Abur et al[18] showed that patients with hyperfunctional voice disorders had poorer PD compared to speakers without a voice disorder. The changes in PD have been considered as possibly playing a role in pathogenesis of behavioral voice disorders. Stepp et al[19] found that patients with hyperfunctional voice disorders showed different adaptive responses in pitch perturbation tasks compared to controls. Muscle Tension Dysphonia patients were found to have a significantly larger magnitude adaptive response to changes to auditory feedback as compared to people without dysphonia, suggesting a possible link between pitch perception and voice production.[20] Pitch pattern recognition abilities as assessed in three-tone sequences were poorer in females with voice disorders in comparison to females without disorders and was also strongly correlated to reduced vocal reproduction of musical tones.[21]

The effects of voice production on auditory perception

There is limited understanding of the effects of voice production on auditory perception. Previous research has shown auditory-perception difficulties in patients with vocal dysfunction, suggesting that impairments in voice production can negatively impact on perception. For example, patients with unilateral vocal fold paralysis (UVFP) showed differences in neural areas associated with vocal-motor function,[22] reduced auditory processing ability, and reduced vocal motor function compared to non-patients, despite receiving adequate surgical treatment.[23] Reduced auditory perceptual abilities were also observed in females diagnosed with behavioral voice disorders with benign vocal fold mucosal lesions,[21] however, the direction of causality of auditory perceptual abilities and voice disorder is not evident in this study. The "*Linked Dual Representation Model*" suggests that vocal production can be mediated by high-level awareness and perception, and low-level (non-conscious) perception.[24] On the one hand, pitch perception abilities would be linked to production through training, such that training in pitch perception transfers to voice production. On the other hand, pitch perception and voice production are thought to be linked through non-aware production pathways. In the latter case, adaptations in voice are made online due to low-level auditory feedback-based corrections.

Some other authors, however, have not observed a significant correlation between PD and voice. Davis and Boone[25] did not find significant differences between the two groups when comparing PD and tonal memory between patients with hyperfunctional voice disorders and control speakers. Murray et al[26] found no relationship between PD and voice production in children with and without vocal nodules.

Measurement of auditory perception

Pitch discrimination (PD) has been assessed in isolation using pitch-based tasks, such as two-tone PD tasks in varying frequency discrimination protocols when comparing trained to untrained musicians and vocalists. Another popular laboratory-based method of measuring pitch perception is known as identifying the just noticeable difference[27] between pitches, or the DLFs.[14] An adaptive tracking procedure is used to obtain a DLF, whereby pitch differences are reduced following a specified number of correct responses and increased followed by one incorrect response.[16,28] Whilst these protocols are successful in identifying the smallest pitch difference that individuals can recognize, the process is long, complex, and the equipment required to conduct these tests is not readily available in clinical settings. As such, two-tone PD tasks may provide the most direct and simple measure of the auditory perceptual system.[15,17,29]

Pitch perception and voice quality

Voice quality has a complex and oft described relationship with signal frequency characteristics and thus can affect perception of pitch. Voice quality descriptors such as roughness, breathiness and hoarseness are associated with acoustic measures such as jitter, which quantify the amount of variations in the fundamental frequency (F0) in prolonged vowel tasks.[30,31] Rough voices and voices characterized by vocal fry are perceived to have a lower pitch than non-disordered voices[32,33] and breathy voices have a higher pitch than non-disordered voices.[32] Given that small variations in fundamental frequency can affect perceptions of voice quality, some acoustic measures of voice quality may be more or less appropriate and possibly more sensitive to fine voice-motor control than others. Similarly, voice quality may also influence perception of pitch. Collecting both voice quality data and auditory perception of PD data in the same population can help to clarify whether such relationships exist and what this might mean, based on musical background, for transfer pathways that are more conscious and explicit. Therefore, we included a range of acoustic voice production measures in our study that relate to not only pitch, but perception of voice quality.

To date, in no study has the relationship between acoustic voice measures and PD been investigated in a large non-clinically disordered cohort where differences between individuals in both voice and PD[34] are apparent but not clinically significant. Specifically, it is unclear whether frequency-domain measures (fundamental frequency and its standard deviation) and spectral-domain measures related to voice quality are correlated with PD accuracy. The aims of the present study were to: 1) Investigate PD characteristics in a cohort of female speakers who were self-identified as non-voice disordered; and 2) Investigate acoustic voice characteristics and their relationship to PD in a cohort of female speakers self-identified as having no voice problems. Overall, we hoped to better evaluate the inherent and/or trained links between pitch perception and various voice production measures to help understand the co-development or otherwise of these skills.

MATERIAL AND METHODS

Ethical approval

The study protocol was approved by The University of Sydney Human Research Ethics Committee (Protocol number: 2016/1001). Written informed consent was obtained from all participants to partake in this study. The study was implemented in accordance with relevant ethical guidelines and regulations.

Participants

There were 114 female participants with a mean age of 23.1 years, standard deviation, SD = 3.8, range = 18-40 years. All were first or second language English-speaking university students. Inclusion criteria included: 1) No self-reported or previous diagnosis of voice disorder; 2) Normal hearing (passing 20-decibel threshold in a pure-tone audiometric screening on frequencies of 1kHz, 2kHz and 4kHz); 3) Non-smokers; 4) Did not regularly use inhaled corticosteroids (a common medication for asthma known to impact voice quality); and 5) Had not experienced upper respiratory problems within 2 weeks before the study.

Participants completed a case history questionnaire to determine history of voice disorders, smoking, upper respiratory problems, language backgrounds, musical background, and voice/musical training. They also completed the Screening Index for Voice Disorders[35] (SIVD) and Voice Handicap Index (VHI-10).[36] Table 1 shows the voice disorder screening results using the SIVD and the VHI-10 scores. All participants had a SIVD score of 3 or below (where 5 or above indicates voice disorder[35]). Although the majority of people had a VHI-10 score below 7.5, there were 15 who had a score above (which is the cut-off value for determination of voice quality handicap).[37]

Participants were classified based on musical background as determined from the case history questionnaire. Musically trained individuals were defined as those having formally learnt a musical instrument for at least a year past the age of 5 years old in individual lessons or in band rehearsals. This resulted in 58 participants defined as musically trained and 56 as non-musically trained. A small proportion of the musicians were identified as having voice training at some point during their developmental years or during adulthood; either individual lessons or in group choir (n = 10). Due to the potential significance of differences between these individuals and instrumentalists (where one group has training in voice and the other does not), secondary comparative analyses were run on these groups.

Pitch discrimination tasks

We used the Newcastle Assessment of Pitch Discrimination (NeAP),[39] which is a two-tone discrimination task. Auditory stimuli were computer-generated tones played on a Dell computer (Optilex 760) via a speaker system (Harman/Kardon HK645) calibrated to 65-65.2 dBA hearing level. Hearing level was measured at ear level using lingWAVES SPL meter II model IEC 651. Participants completed the default protocol that was pre set on the NeAP software. They were instructed to listen to two tones and to indicate which tone was higher in pitch or if they were the same. Responses were provided by clicking on the relevant icon on the computer screen (‘1’, ‘2’ or ‘same’). The default protocol contained twenty pairs of sine waves presented twice for a total of forty tone pairs; tones were presented in random order. Frequencies ranged between B2 (123.47 Hz) to D#4 (311.13 Hz) and pitch differences between tones ranged between one tone and a third of a semitone (Appendix 1). There was an average completion time of ~5 minutes. Participants completed the PD task before completing their voice recordings. The percentages of accurate responses were calculated for each tone pair (t = 20) by dividing the number of accurate responses by the total responses for that tone pair. All tone pairs (total 20) were presented a second time in random order in the same session for reliability analyses.

TABLE 1.
Descriptive Statistics and Cut-off Scores for the SIVD, VHI-10 Scores, Dysphonic Severity Ratings and HNR for the Whole Sample (n = 114)

	Sample Means (SD)	Sample Medians	Sample Range	Cut-off Values (Ref.) & max.
SIVD	0.14 (0.46)	0	0-3	5[35], max = 12
VHI-10	3.51 (3.53)	3	0-15	7.5[37], max = 40
Auditory perceptual severity ratings score	5.02 (3.16)	4.50	0.25-15.75	NA, (max = 100)
HNR	24.05 (3.13)	24.25	16.2-31.2	20[38], max = NA

Note that cut-off value refers to the clinical reference value for determination of a mild voice disorder (along with a publication reference, Ref) and scale maximum (max) values.
Abbreviations: HNR, harmonics-to-noise ratio in dB; NA, not available; SD, standard deviation; SIVD, Screening Index for Voice Disorders; VHI-10, Voice Handicap Index − 10.

Voice recordings

Voice recording took place in a soundproof booth. Participants were fitted with a calibrated, head-mounted, cardiod condenser microphone AKG C520 placed 5 cm and 45° away from the centre of their mouth. The microphone was calibrated using a single sine wave stimulus with frequency of 333.3 Hz at an average intensity of 60 dBA. Voice recordings were made using a Layla 24/96 Multitrack Recording System and Adobe Audition software (Version 1.5)[40] at 44.1 kHz and 16-bit and saved in *.wav format. Participants were required to read the full Rainbow Passage[41] and six sentences of the Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V)[42] and to sustain the vowel 'ah' for 6 seconds. Participants were recommended to use their most comfortable pitch and loudness in producing these vocal tasks. The Rainbow Passage was selected for acoustic analysis as it is a phonemically balanced paragraph.[41] The use of prolonged vowel tasks ensures measurement of vocal perturbation and glottal noise are more accurate than in continuous speech tasks, where these measures are often influenced by intonation and other effects.[43] All vocal tasks (sustained vowel, CAPE-V phrases, and Rainbow Passage) were repeated for a total of three trials and an average of the three trials was used for statistical analysis.

Acoustic voice analysis

Prolonged vowel samples of the vowel 'ah' were trimmed to include two seconds of phonation in the middle section of the voice signal. Samples excluded the first and last one-second region of the signal, as these regions have been reported to contain the highest signal perturbation.[44] Prolonged vowel samples were signal typed using protocols reported by Sprecher et al.[45] Only type 1 and 2 signals were included in the final analysis for frequency-based analyses (such as harmonics-to-noise ratio, HNR) due to their enhanced reliability compared to type 3 and 4 signals.[45]

The 3rd CAPE-V phrase "*We were away a year ago*" (CAPE-V3) was chosen for acoustic analyses based on previous research[46] and its strong correlation to auditory perceptual ratings of voice disorder severity.[47]

The Rainbow Passage[41] voice recordings were edited to include only the second and third sentences "...*The rainbow is a division of white light into many beautiful colours. These take the shape of a long round arch, with its path high above, and its two ends apparently beyond the horizon*...". This task was used to allow comparisons with other studies on cepstral/spectral measures.[48] Connected speech samples were trimmed to have less than 0.5 seconds' silence at the beginning and end of each signal in preparation for acoustic analysis.

Frequency-based acoustic measurements

Frequency-based analysis was performed to obtain acoustic values for F0, standard deviation of F0 (F0SD), HNR, and intensity. These were measured using default settings of the acoustic analysis program Praat version 6.0.25.[49] F0 and F0SD were measured as they are physical acoustic correlates of pitch.[50] HNR was measured as its extraction depends upon reliable identification of pitch boundaries,[51,52] hence it is related to and affected by pitch production. Higher HNR values are correlated with auditory perceptual judgements of better vocal clarity.[38,53] Data for F0 vowel was not included in this study as it does not correspond to F0 in speaking situations.[54] In addition, the F0 used in a prolonged vowel is arbitrary as participants were instructed to sustain the vowel 'ah' at a comfortable pitch for 6 seconds without any verbal model or reference.

Spectral-based measurements

Two spectral-based measures were included in this study: Smoothed Cepstral Peak Prominence (CPPS) and Cepstral/Spectral Index of Dysphonia (CSID). CPPS is obtained from a log power spectrum of a log power spectrum in which its quefrency at the cepstral peak represents the fundamental period of the signal.[55] This measure was included as it is a measure of regularity of the fundamental period with periodic signals showing more prominent cepstral peak (well-defined F0) than aperiodic signals.[56] In addition, CPPS has the strongest correlation to auditory judgements of voice quality in comparison to other acoustic measures.[57,58] Lower CPPS values are correlated with auditory perceptual judgements of poor vocal quality.[59] CPPS values were obtained using SpeechTool.[60]

CSID was included as this measure contains CPP in its formula.[48] CSID was obtained automatically in ADSV Model 5109 Version 3.4.2[61] for vowel (CSIDv) and manually calculated for the Rainbow Passage (CSIDrp) based on CPP, Low/High Spectral Ratio (LH), and standard deviation (SD) of Low/High Spectral Ratio measured in ADSV using the following formula[48]: $\text{CSIDrp} = 154.59 - 10.39{*}\text{CPP} - 1.08{*}\text{LH} - 3.71{*}\text{SD}_{\text{LH}}$

The CSID is a multifactorial measure incorporating the means and standard deviations of Cepstral Peak Prominence and ratio of low versus high frequency spectral energy to provide a quantitative measure for dysphonia.[62] It is reliable and valid in classifying voice disorders.[62] Higher CSID values are correlated with auditory perceptual judgements of poor vocal quality.[62,63] An extended four second prolonged vowel signal was used to calculate CSID due to limitations in the use of the initial two second signal.

Auditory perceptual voice analyses

The auditory perceptual analyses allowed us to rate voice quality of the whole cohort and identify those with potential voice abnormality. The 3 s sustained /a/ vowel and the Rainbow Passage of the second trial were extracted and combined into a separate audio file for each participant. These files were randomized and uploaded onto Bridge2Practice.com[64] for auditory-perceptual assessment. Voice samples were presented in Bridge2Practice.com in random order across raters. Four voice professionals (three speech-language pathologists and one ear, nose and throat specialist doctor) listened to audio samples as many times as they wished and judged the severity of dysphonia by placing a slider on a 100-point rating scale in an online rating form as per the Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V) rating form.[65] An average of the four responses were used for statistical analyses.

Thirty percent of voice samples (35 files) were duplicated, randomized throughout, and rated during the same session to evaluate intra- and inter-rater reliability using intraclass correlation coefficients (ICC). For inter-rater reliability, average ICC was 0.68 (95% CI = 0.56-0.77, $P < 0.01$). Intra-rater reliability for each of the raters ranged from 0.69 to 0.83 (95% CI = 0.36-0.92, $P < .003$). This was considered acceptable for the study as these ICC values would be considered moderate-to-good correlations.[66]

Statistical analysis

Statistical analyses were completed using SPSS 25.0.[67] Data were checked for normal distribution before selecting comparison and correlation tests. When data were normally distributed, subgroup comparisons were made using independent samples *t* tests (Cohen's d is also provided as a measure of effect size). A P value < 0.05 was used to determine statistical significance in group comparison analyses. Pearson's correlation coefficient was used to assess the relationship between individual acoustic voice measures and PD accuracy. Where there were multiple correlation calculations, Bonferroni's adjustment was implemented to minimize Type I error. Bonferroni-adjusted p was deemed significant if it was < 0.05. Receiver Operating Characteristic (ROC) curve analysis was used to calculate the sensitivity and specificity of the NeAP in differentiating musical from non-musical training groups. The significance level was $P < 0.05$ as P values that are too small can lead to Type II errors.

RESULTS

Reliability, sensitivity, and specificity of the NeAP

As the NeAP has not been used clinically, it was important to test the validity of this tool. The intra-class correlation coefficient (ICC) was calculated to assess reliability of agreement across tone pairs, based on the entire cohort. Correlations were all high and positive, both between single measures (ICC = 0.853, 95% CI = 0.794-0.896, $P < 0.001$) and average measures (ICC = 0.921, 95% CI = 0.885-0.945, $P < 0.001$).

ROC curve was calculated from NeAP scores with musical background being the defining variable of positive (n = 58) and negative (n = 56). The NeAP could differentiate the two groups with reasonable balance (highest sensitivity with highest specificity) at a sensitivity of 74.14% and specificity of 73.21%. The optimal cut-off point was identified with Youden J index[68] of 0.474 at PD score ≥ 75%. The area under the ROC curve (AUC) is shown in Figure 1 [95% CI = 0.690-0.850, Z = 6.372, $P < 0.001$] Figure 1. shows that this NeAP (default protocol) had moderate-good prediction accuracy in differentiating musical training background.

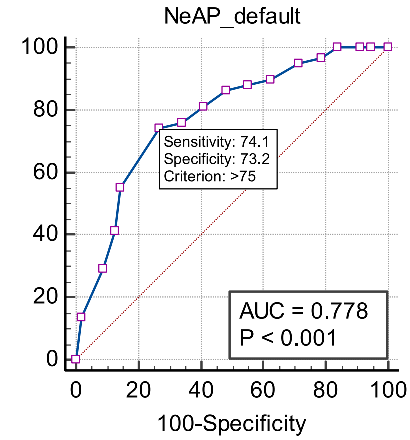

FIGURE 1. ROC curve of percentage of correct response of NeAP. AUC, area under the ROC curve.

Pitch discrimination accuracy

PD accuracy was calculated as the percentage of correct responses from the 20 tone pairs. Although normative data were not currently available for the NeAP test, PD was generally high (above 70%).

Pitch discrimination accuracy in musically trained and non-trained groups

Mean (SD) of PD accuracy (%) was 71.45 (21.11), 81.29 (16.32), and 61.25 (20.76) for the whole sample, musically trained group, and non-musically trained groups, respectively. The musically-trained group had a significantly higher percentage of accurate responses than the non-trained group, $t(104.35) = 5.72$, $P < 0.001$, $d = 1.12$. The maximal pitch difference between tone pairs that were incorrect showed that the non-trained group made errors with larger pitch differences

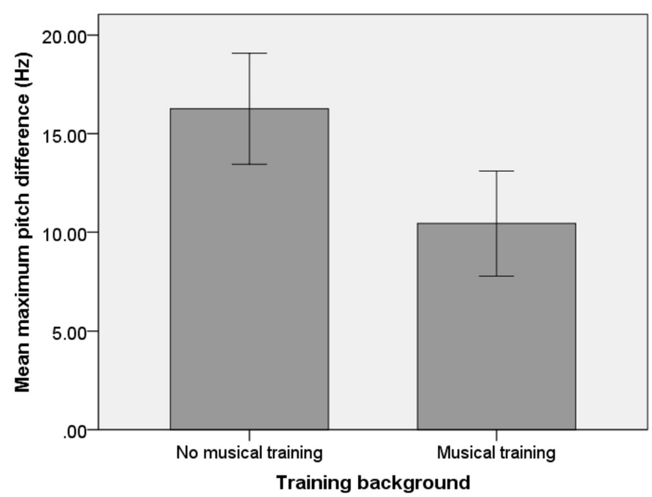

FIGURE 2. Maximal pitch difference of incorrect tone pairs for the musically trained and non-trained group (note error bars show 95% confidence intervals).

TABLE 2.
Sample Means (SD) for Voice Data for the Whole Sample and for the Musically Trained and Non-Musically Trained Groups

Tasks	Acoustic Measures	Whole Sample (n = 114)	Musical Trained Group (n = 58)	Non-Musical Trained Group (n = 56)	*P*-values and Cohen's d	Normative Data (Ref.)
Rainbow passage	F0 (Hz)	201.34 (19.72)	200.26 (14.99)	202.45 (23.74)	0.56; 0.12	171-275[2]
	CPPS (dB)	4.30 (0.54)	4.34 (0.52)	4.26 (0.56)	0.42; 0.15	>4.04[69]
	CSID	-0.83 (18.89)	0.51 (10.12)	-2.21 (24.96)	0.45; 0.18	<19.09[62]
	Intensity (dB)	39.13 (3.57)	38.95 (3.80)	39.32 (3.33)	0.58; 0.10	68.15[2]
3rd CAPE-V phrase	F0 (Hz)	201.61 (19.26)	200.55 (15.65)	202.71 (22.48)	0.55; 0.12	171-275[2]
	CPPS (dB)	5.67 (0.90)	5.70 (0.73)	5.64 (1.06)	0.75; 0.06	NA
	CSID	-15.40 (9.13)	-17.33 (8.12)	-13.41 (9.74)	0.02*; 0.44	NA
	Intensity (dB)	42.18 (3.73)	42.08 (3.92)	42.48 (3.54)	0.78; 0.05	NA
/a/ vowel	F0SD (Hz)	1.43 (0.46)	1.38 (0.42)	1.47 (0.50)	0.31; 0.19	20-29y: 3.8[70] 30-40y: 2.5[71] 40-50y: 2.8[71] 60-69y: 4.3[70]
	HNR (dB)	24.05 (3.13)	24.48 (2.89)	23.61 (3.32)	0.13; 0.28	>20[38]
	CPPS (dB)	8.30 (1.94)	8.42 (2.41)	8.17 (1.30)	0.48; 0.15	>6.12[69]
	CSID	-9,05 (8.79)	-11.02 (7.97)	-7.00 (9.20)	0.01*; 0.47	NA
	Intensity (dB)	43.50 (5.48)	44.03 (5.84)	42.94 (5.06)	0.29; 0.20	NA

Cohen's d and *P*-values are based on (*t* test) comparisons between the musically trained and non-trained groups. References (Ref) are provided to alert to normative data comparisons where available (NA, not available).
(*) significant at $P < 0.05$.
Abbreviations: CPPS, smoothed cepstral peak prominence; CSID, cepstral spectral index of dysphonia; F0, fundamental frequency; HNR, harmonics-to-noise ratio; SD, standard deviation.

than the musically trained group, t(100) = 3.02, P = 0.003, d = 0.60; as shown in Figure 2.[1]

Pitch discrimination accuracy in instrumentalist and vocalist groups

Mean (SD) of PD accuracy (%) for instrumentalist and vocalist group was 79.90 (17.3) and 88.00 (7.89), respectively. The vocalist group had a significantly higher percentage of accurate responses than the instrumentalist group, t (30.27) = 2.23, P = 0.03, d = 0.83. Analysis of the maximal pitch difference between tone pairs that were incorrect showed that the instrumentalist group made errors with larger pitch differences (mean = 12.2, SD = 9.9 Hz) than the vocalist group (mean = 5.3, SD = 1.4 Hz), t (45.527) = 4.305, $P < 0.001$, d = 1.585.

Acoustic voice characteristics

Acoustic voice measures for the whole cohort, along with normative comparative data for the main measures are displayed in Table 2. In general, the sample showed acoustic voice characteristics that were within published normal ranges. Mean CPPS values were higher than the cut-off values for individuals with dysphonia.[69] Mean CSID scores from the current sample were below the cut-off values for dysphonia. Mean HNR for the prolonged vowel was above 20 dB, indicating a normal voice.[38]

Musically trained and non-trained groups

The musically trained group had significantly lower CSID of vowel, t(112) = 2.49, P = 0.01, d = 0.47 and CSID of the 3rd CAPE-V phrase, t(112) = 2.34, P = 0.02, d = 0.44, in comparison to the non-trained group as shown in Table 2. No other acoustic voice measures distinguished the subgroups.

Instrumentalist and vocalist groups

Comparisons across the instrumentalist and vocalist subgroups are shown in Table 3. The vocalist group had significantly lower vowel F0SD, t(56) = 2.05, P = 0.05, d = 0.55 in comparison to the non-trained group. No other voice production measures differentiated the subgroups.

Relationships between pitch discrimination accuracy and acoustic voice measures

Correlations between the measures were first performed on the whole sample as detailed on the left of Table 4. No voice quality measures showed correlations with PD accuracy across the whole cohort (all r values < 0.2).

Within group relationships for the musically trained and untrained speakers

In the musically trained group, a small negative correlation was observed between pitch accuracy and vowel F0SD only, r = -0.37, P = 0.004, with pitch accuracy explaining

[1]Correlations between vocal intensity, CPPS and HNR were analysed in view of the impact of vocal intensity on CPPS and HNR.[74] In the non-trained subgroup, vocal intensity in the Rainbow Passage was correlated with both Rainbow Passage CPPS, r = 0.61 and Vowel HNR, r = 0.46 (both $Ps < 0.001$). No correlations between vocal intensity, CPPS and HNR were observed in the mildly voice disordered and non-disordered subgroups.

TABLE 3.
Sample Means (SD) for Voice Data for the Instrumentalist and Vocalist Groups

Tasks	Acoustic Measures	Instrumentalist Group (n = 48)	Vocalist Group (n = 10)	*P*-values and Cohen's d	Normative Data
Rainbow passage	F0 (Hz)	201.11 (14.07)	196.19 (19.18)	0.35; 0.25	171-275[2]
	CPPS (dB)	4.36 (0.53)	4.27 (0.50)	0.63; 0.13	>4.04[69]
	CSID	0.66 (10.36)	-0.24 (9.34)	0.80; 0.07	<19.09[62]
	Intensity (dB)	39.30 (3.74)	37.31 (3.90)	0.13; 0.41	68.15[2]
3rd CAPE-V phrase	F0 (Hz)	201.51 (13.97)	195.94 (22.43)	0.31; 0.27	171-275[2]
	CPPS (dB)	5.67 (0.71)	5.81 (0.82)	0.59; 0.14	NA
	CSID	-17.11 (8.12)	-18.39 (8.47)	0.65; 0.12	NA
	Intensity (dB)	42.39 (3.68)	40.62 (4.89)	0.20; 0.35	NA
/a/ vowel	F0SD (Hz)	1.43 (0.42)	1.14 (0.36)	0.05*; 0.55	NA
	HNR (dB)	24.47 (2.81)	24.54 (3.44)	0.95; 0.02	>20dB[38]
	CPPS (dB)	8.53 (2.50)	7.93 (1.95)	0.48; 0.19	>6.12[69]
	CSID	-10.99 (7.39)	-11.13 (10.82)	0.96; 0.01	NA
	Intensity (dB)	42.34 (6.65)	43.50 (5.48)	0.32; 0.27	NA

Cohen's d and *P*-values are based on (*t* test) comparisons between the instrumentalist and vocalist groups. References (Ref) are provided to alert to normative data comparisons where available (NA, not available).
(*) significant at $P \leq 0.05$.
Abbreviations: CPPS, smoothed cepstral peak prominence; CSID, cepstral spectral index of dysphonia; F0, fundamental frequency; HNR, harmonics-to-noise ratio; NA, not available; SD, standard deviation.

TABLE 4.
Pearson's Correlation Coefficients (r) for the Pitch Discrimination and Voice Data for the Whole Sample, Musically Trained and Non-Musically Trained Groups and Healthy Voice and Mildly Voice Disordered Groups

Tasks	Acoustic Measures	Whole Sample (n = 114)	Musical Trained Group (n = 58)	Non-Musical Trained Group (n = 56)	Instrumentalists (n = 48)	Vocalists (n = 10)
Rainbow passage	F0 (Hz)	-0.10	0.11	-0.19	0.12	0.35
	CPPS (dB)	-0.08	0.13	**-0.33***	0.12	0.38
	CSID	0.10	-0.17	0.15	-0.20	0.21
	Intensity (dB)	-0.11	0.14	-0.32	0.17	0.46
3rd CAPE-V phrase	F0 (Hz)	-0.20	-0.09	-0.26	-0.11	0.16
	CPPS (dB)	0.03	0.16	-0.06	0.14	0.30
	CSID	-0.11	-0.13	0.09	-0.09	-0.51
	Intensity (dB)	-0.06	0.15	-0.24	0.19	0.30
/a/ vowel	F0SD (Hz)	-0.18	**-0.37***	0.01	**-0.39***	0.29
	HNR (dB)	-0.07	0.02	-0.29	0.02	-0.05
	CPPS (dB)	0.05	-0.06	0.13	-0.05	0.11
	CSID	-0.18	-0.10	-0.08	-0.15	0.24
	Intensity (dB)	-0.02	0.04	-0.20	0.04	0.36

All measures are in dB, unless otherwise stated.
(*), Bonferroni-adjusted $P < 0.05$.
Abbreviations: CPPS, smoothed cepstral peak prominence; CSID, cepstral spectral index of dysphonia; F0, fundamental frequency; HNR, harmonics-to-noise ratio; SD, standard deviation.

about 14% of the variance in this measure. Greater PD was associated with low variance in fundamental frequency production. We have plotted data for the whole sample in Figure 3, with the individuals with a musical background shown as solid symbols. No other acoustic measures showed correlations with PD accuracy (all r values < 0.3).

In the non-trained subgroup (n = 56), small negative correlations were observed between pitch accuracy and Rainbow Passage Intensity (r = -0.32, P = 0.02) and Rainbow Passage CPPS (r = -0.33, P = 0.01). Better PD was also associated with worse performance on the vowel HNR voice measures (r = -0.29, P = 0.03), although this correlation was not significant at the adjusted p level. Other acoustic measures did not show correlations with pitch accuracy (all r values < 0.30).[1]

[1]Correlations between vocal intensity, CPPS and HNR were analysed in view of the impact of vocal intensity on CPPS and HNR.[74] In the non-trained subgroup, vocal intensity in the Rainbow Passage was correlated with both Rainbow Passage CPPS, r = 0.61 and Vowel HNR, r = 0.46 (both Ps < 0.001). No correlations between vocal intensity, CPPS and HNR were observed in the mildly voice disordered and non-disordered subgroups.

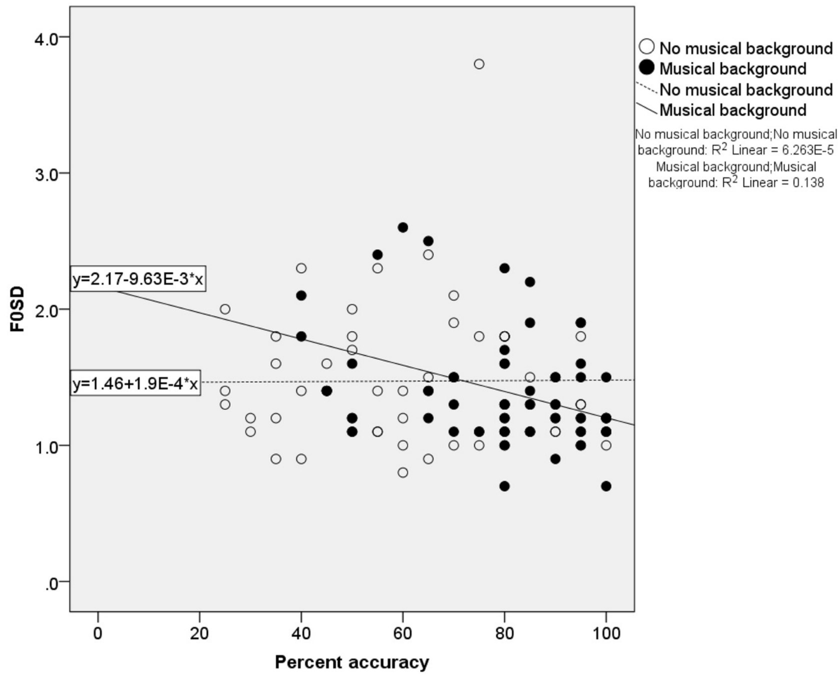

FIGURE 3. Scatter plot depicting the relationship between perceptual discrimination accuracy (%) and F0SD voice quality for musically trained (solid symbols/regression line) and non-trained participants (open symbols/dashed regression line).

Within group relationships for the instrumentalist and vocalist groups

Among the relatively large group identified as instrumentalists, a small correlation was observed between PD accuracy and F0SD (r = -0.39, P = 0.006). Among the vocalists, there were a number of correlations of 0.30 or greater between PD accuracy and acoustic voice measures, but due to the small sample size, none of these were statistically significant. Of most note was the correlation of -0.51 between PD accuracy and CAPE-V3 CSID, meaning that better PD accuracy was associated with better voice quality, however this correlation was not statistically significant (Table 4).

DISCUSSION

There were no statistically significant relationships between any of our various measures of voice quality and pitch perception among our sample of 114 female participants who self-reported as having a non-disordered voice. Individual differences in perception of pitch were not concomitantly manifest in individual differences in voice production. Although musicians had generally better pitch perception than non-musicians, there was little evidence for a relationship between production and perception. Moreover, the musicians only differed from the non-musicians in one out of 13 acoustic voice measures.

Validity of NeAP as a pitch discrimination testing tool

The NeAP was shown to be a reliable tool in testing PD ability. It showed moderate to good prediction accuracy in differentiating musical training background. Whilst the PD accuracy in this study was similar to other research in non-disordered speakers,[17] further studies to establish the best protocol to use in terms of number of tone pairs and size of tone difference should also be explored. It is possible that with measures of just noticeable difference, there may be increased variance and sensitivity to pitch differences.

Pitch discrimination accuracy and voice characteristics

The whole cohort had a PD accuracy of approximately 70%. Currently, no large cohort baseline data exists on non-clinical populations for voice, so these data on pitch perception will help to establish such norms. This data is comparable to the criteria of 75% discrimination accuracy used to distinguish participants who ‘accurately discriminated’ tone pairs in a previous study of singers, musicians, and non-musicians.[17]

Pitch perception and voice quality in musically trained and untrained speakers

Consistent with previous studies,[14,16,27,72] musical training background had a strong effect on PD accuracy, even though none of the participants identified as professional musicians. Participants with musical training had better PD accuracy than those without. In the musically trained subgroup, PD accuracy was only correlated with 1 out of 13 acoustic voice measures. Higher PD accuracy was weakly correlated with lower F0SD, which is indicative of better control of F0 in a prolonged vowel task. This general absence of a relationship between PD and acoustic voice data may be a consequence of the fact that participants had low, variable and for the most part, early development musical training. Indeed, positive correlations were observed between PD ability and pitch matching accuracy in musicians in a previous study.[14] The use of different measures of PD and vocal production across studies might partially explain the lack of any significant relationship.

We used a range of acoustic voice measures, including both frequency-based and spectral-based, to represent overall vocal function/production. Whilst the frequency measures (F0 and F0SD) were more relevant to the PD task, other acoustic measures reflect the overall voice quality (ie, CPPS and CSID). The control of voice production requires more than just the control of pitch. For participants with musical training experience, the auditory system would likely have been trained to attend to pitch, such as when tuning an instrument. As such, there is reason to think it would also be more sensitive to minor pitch changes in voice. For those without a musical background, the mechanism of voice control may not be driven by factors related to pitch. Therefore, the use of overall voice quality measures was reasonable. Our results seemed to suggest that conscious, auditory perception might play a weak role compared with other sensory feedback modalities such as proprioception, or non-conscious perceptual processes as suggested in the Linked Dual Representation Model [23], in regulation of vocal production.

In the musically untrained subgroup, PD accuracy was weakly correlated with 1 out of 13 acoustic voice measures and another two vocal production measures showed trends of correlation with PD accuracy. Individuals with better PD accuracy were found to havea lower CPPS of the Rainbow Passage and lower HNR in the prolonged vowel task, indicative of a poorer voice quality. Existing evidence at most shows a dissociation between vocal perception and production,[17,73] not a negative relationship. It may be that the lower CPPS values were caused by reduced vocal intensity. In the non-trained group, PD score showed a trend to be correlated with vocal intensity of the Rainbow Passage (r = -0.32, Table 4). This is consistent with existing evidence that shows lower vocal intensity is correlated with lower CPPS and HNR values[1].[74] As such, it is likely that these weak associations were a consequence of using a softer voice rather than a degraded voice quality. It is possible that untrained speakers use a different mechanism to control their voices compared to musically trained people, monitoring their vocal loudness rather than pitch. Indeed, in a typical speaker, vocal intensity is an important feature that ensures communicative effectiveness, especially when speaking in an environment with ambient noise.[75] It has been shown experimentally that when speakers are trained to produce a novel vocal task, individuals make significant changes to reduce vocal intensity to achieve improved vocal productions.[76,77]

Pitch perception and voice quality in instrumentalists and vocalists

In contrast to previous studies,[14,16] vocalists had significantly better PD abilities and more stable fundamental frequency control than instrumentalists (although there were only n = 10 vocalists causing statistical issues in variance comparing across vocalists and instrumentalists, of which the latter group had n = 48). Whilst vocalists did not demonstrate better vocal production abilities in other acoustic measures, better PD supports the idea that training effects are more specific rather than general. For participants with vocal training experience, the vocal production system would likely have been trained to match or hold a stable pitch when singing, therefore, it is likely that vocalists would be more sensitive to minor pitch changes. As such, our results suggest that vocal production in specific tasks may play a small role in regulating auditory perception abilities.

PD accuracy only showed a trend to be correlated with more stable fundamental frequency control (lower F0SD) in instrumentalists. No statistically significant correlations were observed between PD accuracy and vocal production measures in vocalists (although there were small to medium sized effects noted for a number of measures and all but three of the correlations were greater than r = 0.20, Table 4). The lack of significant correlations observed in this study is consistent with results in highly trained, working vocalists,[14] and may be due to low variability in both PD abilities and vocal production in this group.

With the extremely small sample of vocalists (n = 10), any conclusions based on comparisons or correlations should be treated cautiously. In future work it will be important to consider a larger cohort of vocalists to understand the relationship between vocal perception and production abilities; ideally performing developmental work to determine if and how these abilities co-develop.

Implications for theoretical models of perception and production

The findings can be used to understand some theoretical models related to perception and production. According to the Theory of Event Coding,[78,79] we would expect correlations between perception and production in a non-disordered population of speakers. However, PD was not significantly correlated with any of the vocal production measures in the general cohort. Correlations with PD accuracy were only observed in participants with musical training background and only in measures closely associated to specific skills (ie, pitch production) that is important in musical training.

[1]Correlations between vocal intensity, CPPS and HNR were analysed in view of the impact of vocal intensity on CPPS and HNR.[74] In the non-trained subgroup, vocal intensity in the Rainbow Passage was correlated with both Rainbow Passage CPPS, r = 0.61 and Vowel HNR, r = 0.46 (both Ps < 0.001). No correlations between vocal intensity, CPPS and HNR were observed in the mildly voice disordered and non-disordered subgroups.

Based on the Linked Dual Representation[24] model, pitch perception and vocal production can be modulated in two separate pathways. In one pathway, vocal production can be mediated by conscious perceptual judgments of sound, whereby some sort of perceptual trace or representation guides action production.[24] Accordingly, those with better pitch perception abilities will often also have better pitch production abilities, but the reverse relationship does not hold. Our data is not consistent with this uni-directional relationship between perception and production. We did not observe a relationship between PD and any acoustic voice measures in the whole cohort. In addition, it was the vocalists who had more stable vocal production skills that were linked to better PD, compared to instrumentalists. This vocalist-instrumentalist discrepancy suggests a bi-directional relationship between voice production and PD. Alternatively, people with naturally stable phonation may tend to become vocalists rather than instrumentalists. They may also have better pre training PD compared to instrumentalists and musically untrained people; however, this should be investigated in future studies.

CONCLUSION

The relationship between auditory perception and vocal production continues to be elusive, yet important for theoretical and practical reasons. Based on multiple measures of acoustic voice quality in this study, it would appear that in a population of individuals who would be considered clinically non-voice disordered, with low musical training, there is little to no relationship between PD and voice quality. Where relationships were observed, these were suggestive of a uni-directional relationship whereby conscious perception influences vocal production in individuals with some musical training, consistent with the Linked Dual Representation[24] model. These perceptual to production transfer benefits were limited to measures specific to pitch, in this case fine control of fundamental frequency. These results support the idea that training the perceptual system will influence the control of the production system in the specific domain in which it is trained (eg, frequency, intensity or spectral features). However, the reverse relationship was not observed in this study. Although vocalists had improved pitch perception skills compared to instrumentalists, we did not observe statistically significant correlations between perception and production measures (likely because we were underpowered to detect such effects). These results suggest that we cannot rule out a bi-directional relationship between perception and production. Our data also provide preliminary evidence that individuals without musical training might monitor their vocal production through the perception of intensity (ie, loudness). Future studies must therefore ensure that the PD and acoustic voice measures used are appropriate to the skills that are being investigated.

AUTHOR CONTRIBUTION STATEMENTS

Emily Wing-Tung Yun conducted the literature review, prepared the research protocol, recruited participants, collected data, performed acoustic voice analyses and wrote the manuscript. Duy Duong Nguyen was involved in research question identification, data collection, data analysis, data interpretation, manuscript writing and editing, and graphic works. Paul Carding and Nicola Hodges were involved with reviewing and editing the manuscript. Robert Heard assisted with data interpretation and analysis. Antonia Chacon was involved in recruitment of participants and data collection. Catherine Madill conceived the research idea, wrote and edited the manuscript. All authors reviewed and approved the final version of the manuscript.

COMPETING INTERESTS

The authors have no competing interests to declare in this study.

Acknowledgments

We'd like to acknowledge the contribution of Dr. Robert Heard at The University of Sydney for his initial advice regarding statistical analysis and research design.

APPENDIX 1. REFERENCE AND COMPARISON TONES USED IN THE DEFAULT NEAP PD TASK

	Frequency of Tones (Hz)	
Trial	Reference Tone	Comparison Tone
1	B2 (123.47)	C3 (130.81)
2	C3 (130.81)	D3 (146.83)
3	C3 (130.81)	C#3 (138.59)
4	C3 (130.81)	C3.5 (134.64)
5	C3 (130.81)	C3.3 (133.1)
6	D3 (146.83)	D3.5 (151.13)
7	E3 (164.81)	F3 (174.61)
8	E3 (164.81)	E3.5 (169.64)
9	F3 (174.61)	F#3 (185.00)
10	F3 (174.61)	F3.5 (179.73)
11	F3 (174.61)	F3.3 (177.66)
12	G3 (196.00)	G#3 (207.65)
13	G3 (196.00)	G3.5 (201.74)
14	A3 (220.00)	B3 (246.94)
15	A3 (220.00)	A#3 (233.08)
16	A3 (220.00)	A3.5 (226.45)
17	B3 (246.94)	B3.5 (254.18)
18	C4 (261.63)	D4 (293.66)
19	D4 (293.66)	D#4 (311.13)
20	D4 (293.66)	D4.5 (302.26)

REFERENCES

1. Heller Murray ES, Stepp CE. Relationships between vocal pitch perception and production: a developmental perspective. *Sci Rep*. 2020;10:1–10.
2. Colton RH, Leonard R, Casper JK. *Understanding Voice Problems: A Physiological Perspective for Diagnosis and Treatment*. Philadelphia, Pa: Lippincott Williams & Wilkins; 2011.

3. Beck SL, Rieser JJ, Erdemir A. Singing without hearing: a comparative study of children and adults singing a familiar tune. *Psychomusicology*. 2017;27:122–131.
4. Bergan C. Mindful Voice - Motor learning principles and voice pedagogy: theory and practice. *J Singing*. 2010;66:457–468.
5. Bottalico P, Graetzer S, Hunter EJ. Effect of training and level of external auditory feedback on the singing voice: pitch inaccuracy. *J Voice*. 2017;31. 122.e9-.e16.
6. Abur D, Daliri A, Stepp CE, et al. Evidence for auditory-motor impairment in individuals with hyperfunctional voice disorders. *J Speech Lang Hear Res*. 2017;60:1545–1550.
7. Jones J, Keough D. Auditory-motor mapping for pitch control in singers and nonsingers. *Exp Brain Res*. 2008;190:279–287.
8. Burnett TA, Senner JE, Larson CR. Voice F_0 responses to pitch-shifted auditory feedback: a preliminary study. *J Voice*. 1997;11:202–211.
9. Smith DJ, Stepp C, Guenther FH, et al. Contributions of auditory and somatosensory feedback to vocal motor control. *J Speech Lang Hear Res*. 2020;63(7):2039–2053.
10. Bradshaw E, McHenry MA. Pitch discrimination and pitch matching abilities of adults who sing inaccurately. *J Voice*. 2005;19:431–439.
11. Kishon-Rabin L, Amir O, Vexler Y, et al. Pitch discrimination: are professional musicians better than non-musicians? *J Basic Clin Physiol Pharmacol*. 2001;12(2 Suppl):125–143.
12. Monsen RB. Voice quality and speech intelligibility among deaf children. *Am Ann Deaf*. 1983;128:12–19.
13. Higgins MB, Carney AE, Schulte L. Physiological assessment of speech and voice production of adults with hearing loss. *J Speech Hear Res*. 1994;37:510–521.
14. Nikjeh DA, Lister JJ, Frisch SA. The relationship between pitch discrimination and vocal production: comparison of vocal and instrumental musicians. *J Acoust Soc Am*. 2009;125:328–338.
15. Estis JM, Dean-Claytor A, Moore RE, et al. Pitch-matching accuracy in trained singers and untrained individuals: the impact of musical interference and noise. *J Voice*. 2011;25:173–180.
16. Nikjeh DA, Lister JJ, Frisch SA. Hearing of note: an electrophysiologic and psychoacoustic comparison of pitch discrimination between vocal and instrumental musicians. *Psychophysiology*. 2008;45:994–1007.
17. Bradshaw E, McHenry MA. Pitch discrimination and pitch matching abilities of adults who sing inaccurately. *J Voice*. 2005;19:431–439.
18. Abur D, Subaciute A, Kapsner-Smith M, et al. Impaired auditory discrimination and auditory-motor integration in hyperfunctional voice disorders. *Sci Rep*. 2021;11:1–11.
19. Stepp CE, Lester-Smith RA, Abur D, et al. Evidence for auditory-motor impairment in individuals with hyperfunctional voice disorders. *J Speech Lang Hear Res*. 2017;60:1545–1550.
20. Ziethe A, Petermann S, Hoppe U, et al. Control of fundamental frequency in dysphonic patients during phonation and speech. *J Voice*. 2019;33:851–859.
21. Ramos JS, Feniman MR, Gielow I, et al. Correlation between voice and auditory processing. *J Voice*. 2018;32(6):771.e25–771.e36.
22. Naunheim ML, Yung KC, Schneider SL, et al. Cortical networks for speech motor control in unilateral vocal fold paralysis. *Laryngoscope*. 2019;129(9):2125–2130.
23. Naunheim ML, Yung KC, Schneider SL, et al. Vocal motor control and central auditory impairments in unilateral vocal fold paralysis: motor and auditory impairments in UVFP. *Laryngoscope*. 2019;129(9):2112–2117.
24. Hutchins S, Moreno S. The linked dual representation model of vocal perception and production. *Frontiers Psychol*. 2013;4:1–12.
25. Davis DS, Boone DR. Pitch discrimination and tonal memory abilities in adult voice patients. *J Speech Hear Res*. 1967;10:811–815.
26. Murray ESH, Hseu AF, Nuss RC, et al. Vocal pitch discrimination in children with and without vocal fold nodules. *Appl Sci (Basel)*. 2019;9:1–13.
27. Arndt C, Schlemmer K, van Der Meer E. Same or different pitch? Effects of musical expertise, pitch difference, and auditory task on the pitch discrimination ability of musicians and non-musicians. *Exp Brain Res*. 2019;238:247–258.
28. Micheyl C, Divis K, Wrobleski D, et al. Does fundamental-frequency discrimination measure virtual pitch discrimination? *J Acoust Soc Am*. 2010;128:1930–1942.
29. D'Ausilio A, Bufalari I, Salmas P, et al. Vocal pitch discrimination in the motor system. *Brain Lang*. 2011;118:9–14.
30. Laver J, Hiller S, Beck JM. Acoustic waveform perturbations and voice disorders. *J Voice*. 1992;6:115–126.
31. Maryn Y, Roy N, Bodt Md, et al. Acoustic measurement of overall voice quality: a meta-analysis. *J Acoust Soc Am*. 2009;126:2619–2634.
32. Munoz J, Mendoza E, Fresneda MD, et al. Acoustic and perceptual indicators of normal and pathological voice. *Folia Phoniatr Logop*. 2003;55:102–114.
33. Kuang J, Liberman M. The effect of vocal fry on pitch perception. *Int Conf Acoust Speech Signal Process; 2016*. Shanghai, China: IEEE; 2016:5260–5264.
34. Smith L, Bartholomew A, Burnham L, et al. Factors affecting pitch discrimination performance in a cohort of extensively phenotyped healthy volunteers. *Sci Rep*. 2017;7:1–9.
35. Ghirardi ACdAM, Ferreira LP, Giannini SPP, et al. Screening index for voice disorder (SIVD): development and validation. *J Voice*. 2013;27:195–200.
36. Rosen CA, Lee AS, Osborne J, et al. Development and validation of the Voice Handicap Index-10. *Laryngoscope*. 2004;114:1549–1556.
37. Behlau M, Madazio G, Moreti F, et al. Efficiency and cutoff values of self-assessment instruments on the impact of a voice problem. *J Voice*. 2015;30. 506.e9-.e18.
38. Warhurst S, Madill C, McCabe P, et al. The vocal clarity of female speech-language pathology students: an exploratory study. *J Voice*. 2012;26:63–68.
39. Drinnan M. Newcastle Assessment of Pitch Discrimination: User Manual 2012 Available at: http://drinnan.net/Site/NeAP_files/Newcastle%20Assessment%20of%20Pitch%20Discrimination%202012-07-11_2.pdf. Accessed March 12, 2018.
40. Adobe Systems Inc. Adobe Audition CC 2018 Available at: https://www.adobe.com/au/products/audition.html?sdid=V6NZKW5P&mv=search&ef_id=WjoC_gAAAHySFHNG:20180516063911:s. Accessed May 28, 2018.
41. Fairbanks G. *Voice and Articulation Drillbook*. 2nd ed. New York: Harper & Row; 1960.
42. Kempster GB, Gerratt BR, Verdolini Abbott K, et al. Consensus auditory-perceptual evaluation of voice: development of a standardized clinical protocol. *Am J Speech Lang Pathol*. 2009;18:124–132.
43. Parsa V, Jamieson DG. Acoustic discrimination of pathological voice: sustained vowels versus continuous speech. *J Speech Lang Hear Res*. 2001;44:327–339.
44. Choi SH, Lee J, Sprecher AJ, et al. The effect of segment selection on acoustic analysis. *J Voice*. 2012;26:1–7.
45. Sprecher A, Olszewski A, Jiang JJ, et al. Updating signal typing in voice: addition of type 4 signals. *J Acoust Soc Am*. 2010;127:3710–3716.
46. Watts CR. The effect of CAPE-V sentences on cepstral/spectral acoustic measures in dysphonic speakers. *Folia Phoniatr Logop*. 2015;67:15–20.
47. Awan SN, Roy N, Jetté ME, et al. Quantifying dysphonia severity using a spectral cepstral-based acoustic index: Comparisons with auditory-perceptual judgements from the CAPE-V. *Clin Linguist Phon*. 2010;24:742–758.
48. Awan SN, Roy N, Zhang D, et al. Validation of the Cepstral Spectral Index of Dysphonia (CSID) as a screening tool for voice disorders: development of clinical cutoff scores. *J Voice*. 2016;30:130–144.
49. Boersma P, Weenink D. Praat: doing phonetics by computer [Computer program]. 6.0.25 ed2017.
50. Baken RJ, Orlikoff RF. *Clinical measurement of speech and voice*. 2nd ed. San Diego: Singular Thomson Learning; 2000.
51. Yumoto E, Gould WJ, Baer T. Harmonics-to-noise ratio as an index of the degree of hoarseness. *J Acoust Soc Am*. 1982;71:1544–1550.

52. Awan SN, Frenkel ML. Improvements in estimating the harmonics-to-noise ratio of the voice. *J Voice*. 1994;8:255–262.
53. Freitas SV, Pestana PM, Almeida V, et al. Integrating voice evaluation: correlation between acoustic and audio-perceptual measures. *J Voice*. 2015;29. 390.E1-.e7.
54. Iwarsson J, Hollen Nielsen R, Næs J. Mean fundamental frequency in connected speech and sustained vowel with and without a sentence-frame. *Logoped Phoniatr Vocol*. 2020;45:91–96.
55. Hillenbrand J, Houde RA. Acoustic correlates of breathy vocal quality: dysphonic voices and continuous speech. *J Speech Hear Res*. 1996;39:311–321.
56. Watts CR, Awan SN, Maryn Y. A comparison of cepstral peak prominence measures from two acoustic analysis programs. *J Voice*. 2017;31. 387 e1- e10.
57. Sauder C, Bretl M, Eadie T. Predicting voice disorder status from smoothed measures of Cepstral Peak Prominence using Praat and Analysis of Dysphonia in Speech and Voice (ADSV). *J Voice*. 2017;31(5):557–566.
58. Maryn Y, Weenink D. Objective dysphonia measures in the program praat: smoothed cepstral peak prominence and acoustic voice quality index. *J Voice*. 2015;29:35–43.
59. Heman-Ackah YD. Quantifying the Cepstral Peak Prominence, a measure of dysphonia. *J Voice*. 2014;28:783–788.
60. Hillenbrand JM. SpeechTool. 2002. Available from: https://homepages.wmich.edu/~hillenbr/. Accessed May 24, 2018.
61. KayPENTAX. *Analysis of Dysphonia in Speech and Voice*. 3.4.2 ed. Montvale, NJ: KayPENTAX; 2011.
62. Awan SN. Validation of the Cepstral Spectral Index of Dysphonia (CSID) as a screening tool for voice disorders: development of clinical cutoff scores. *J Voice*. 2016;30:130–144.
63. Lowell SY, Kelley RT, Awan SN, et al. Spectral- and cepstral-based acoustic features of dysphonic, strained voice quality. *Ann Otol Rhinol Laryngol*. 2012;121:539–548.
64. Madill C, Corcoran S, So T. Bridge2Practice.com 2019 [1:] Available at: https://bridge2practice.com/. Accessed August 8, 2019.
65. Kempster GB, Gerratt BR, Verdolini Abbott K, et al. Consensus auditory-perceptual evaluation of voice: development of a standardized clinical protocol. *Am J Speech-Language Pathol*. 2009;18:124–132.
66. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med*. 2016;15:155–163.
67. Corp. I. *IBM SPSS Statistics for Windows*. 24.0 ed. Armonk, NY: IBM Corp; 2016.
68. Youden WJ. Index for rating diagnostic tests. *Cancer*. 1950;3:32–35.
69. Madill C, Nguyen D, Eastwood C, et al. Comparison of cepstral peak prominence measures using the ADSV, SpeechTool, and VoiceSauce acoustic analysis programs in vocally healthy female speakers. *Acoustics Australia*. 2018;46:215–226.
70. Stoicheff ML. Speaking fundamental frequency characteristics of nonsmoking female adults. *J Speech Hear Res*. 1981;24:437–441.
71. Saxman JH, Burk KW. Speaking fundamental frequency characteristics of middle-aged females. *Folia Phoniatr (Basel)*. 1967;19:167–172.
72. Kishon-Rabin L, Amir O, Vexler Y, et al. Pitch discrimination: are professional musicians better than non-musicians? *J Basic Clin Physiol Pharmacol*. 2001;12:125–144.
73. Amir O, Amir N, Kishon-Rabin L. The effect of superior auditory skills on vocal accuracy. *J Acoust Soc Am*. 2003;113:1102–1108.
74. Sampaio M, Vaz Masson ML, de Paula Soares MF, et al. Effects of fundamental frequency, vocal intensity, sample duration, and vowel context in cepstral and spectral measures of dysphonic voices. *J. Speech Lang Hear Res*. 2020;63:1326–1339.
75. Hilger AI, Levant S, Kim J, et al. Auditory feedback control of vocal intensity during speech and sustained-vowel production. *J Acoust Soc Am*. 2019;146:3052.
76. Joscelyne-May C, Madill CJ, Thorpe W, et al. The effect of clinician feedback type on the acquisition of a vocal siren. *Folia Phoniatrica Et Logopaedica*. 2016;67:57–67.
77. Look C, McCabe P, Heard R, et al. Show and tell: video modeling and instruction without feedback improves performance but is not sufficient for retention of a complex voice motor skill. *J Voice*. 2019;33:239–249.
78. Hommel B, Msseler J, Aschersleben G, et al. The Theory of Event Coding (TEC): a framework for perception and action planning. *Behav Brain Sci*. 2001;24:849–878.
79. Hommel B. Action control according to TEC (theory of event coding). *Psychol Res*. 2009;73:512–526.