Cultural Competency in Voice Evaluation: Considerations of Normative Standards for Sociolinguistically Diverse Voices

*Teresa Procter, and †Ashwini Joshi, *†*Houston, Texas*

Summary: Aim. Significant differences exist in anatomical, acoustic, and aerodynamic parameters for non-speech tasks between culturally and linguistically diverse sample populations. There is a need for expansion of the normative acoustic data to include sociolinguistically diverse groups to ensure that clinical objective measurements are accurately classifying the voice quality of all individuals. This study examined objective measures of voice quality assessment of monolingual speakers of Standard American English (SAE) with sequential bilingual, native (L1) French and Spanish speakers on perturbation, noise, spectral/cepstral analyses, and compared ratings on auditory-perceptual assessment with acoustic data secondary to degree of accentedness.

Method. Thirty speakers with normal voice quality were rated on the Consensus Auditory-Perceptual Evaluation of Voice scale. Voice quality measures were analyzed using the Multi-Dimensional Voice Program and Analysis of Dysphonia in Speech and Voice. A measure of accentedness of SAE was calculated using an informal task by two evaluators.

Results. Objective acoustic measures of jitter and all-voiced cepstral peak prominence were statistically significant between SAE speakers and L1 Spanish and French speakers. SAE speakers demonstrated significantly higher group mean cepstral peak prominence for the all-voiced sentence ("We were away a year ago.") than native French and Spanish speakers. There were no significant differences in perception of voice quality and acoustic measures secondary to degree of accentedness of the non-native SAE speakers.

Conclusion. It is important to engage and strengthen voice diagnostic measures to support culturally competent service delivery for the diversifying clinical population. Normative databases established on SAE speakers should reflect the statistically significant differences evidenced between diverse sociolinguistic populations in anatomical, auditory-perceptual, aerodynamic, and acoustical parameters.

Key Words: Voice evaluation−Cultural competency−Diverse population−Cepstral analysis−Cepstral peak prominence−Accent.

INTRODUCTION

Utilizing cultural competence in clinical practice is a necessary topic in the field of speech-language pathology as the clinical population diversifies. Cultural competence, as defined by the Code of Ethics from the American Speech-Hearing and Language Association (ASHA)[1] is pertinent in service delivery for speech-language pathologists (SLPs) and audiologists. This code asks SLPs to ensure that cultural variables, language exposure, and culturally linguistic differences are taken into consideration when providing care to patients and clients. Data from the 2011 to 2015 United States (US) Census Bureau revealed that 22% of the US population spoke another language at home other than English.[2] US states bordering Mexico had a higher than national average percentage of speaking a language besides English at home (ie, 28% in Texas and 26% in California).[2] Inclusive of all linguistically and culturally diverse individuals, at least one third of the US population may encounter vocal impairment at some point in their lives.[3] Cultural diversity in our clinical population is of salient consideration for diagnostic services. Assessment of voice must therefore be composed of international standardized measures and awareness of language and cultural contributions of voice quality, to ensure high sensitivity in evaluation of all voices.

Accepted for publication September 29, 2020.

From the *Texas Voice Center, Houston Methodist ENT Specialists, Houston, Texas; and the †University of Houston, Houston, Texas.

Address correspondence and reprint requests to Ashwini Joshi, Department of Communication Sciences and Disorders, University of Houston, 4455 Cullen Blvd, Houston, TX 77204. E-mail: ajoshi4@uh.edu

Journal of Voice, Vol. 36, No. 6, pp. 793−801
0892-1997

https://doi.org/10.1016/j.jvoice.2020.09.025

Language: vocal expression through speech

Language is considered a salient tenet of culture. Within the discipline of speech-language pathology, language is a set of universal symbols, built on a structure of phonological and phonetic characteristics used as a communication system to convey and receive messages.[4] While language may be verbal or nonverbal, a verbalized signal through speech engages a corticomotor pathway to activate neurological, respiratory, phonatory, resonatory, and articulatory systems of an individual speaker. Oscillating vocal folds are set in motion by air pressure exchange from the lungs for phonation in all individuals. A signal becomes idiosyncratic to each individual due to the manipulation and interaction of supraglottic anatomical structures that comprise the vocal tract filter (ie, pharyngeal, oral, and nasal cavities), acoustical variations to the message (eg, context, stress, and audience of the message), and speaker-to-speaker differences in the size and shape of the glottis, vocal tract, and the laryngeal structure.[5,6] These differences in physiological, phonological, and phonetic characteristics to produce each language begets the notion that voice is a cultural construct. Therefore, the level of competency in voice diagnostic services for sociolinguistically diverse patients is essential to evaluate.

Anatomical differences in vocal tracts

Anatomically, cross-sectional differences between racial/ethnic diverse populations have been identified in vocal tract parameters and nasal cross-sectional areas. Through the use of acoustic reflection technology in a study of 120 Caucasian-American, African-American, and Chinese female and male speakers, Xue and Hao revealed statistically significant sex and race main effects for all dependent variables in five vocal tract dimensions: oral length, oral volume, pharyngeal length, pharyngeal volume, and vocal tract length.[7] Normative data for these diverse populations were established. Acoustic rhinometry via acoustic reflection also demonstrated statistically significant differences in internal nasal diameters and volumes in the nasopharynx, between four racial/ethnic groups and established normative data.[8,9]

Voice evaluation

Current standards in voice evaluation include objective and subjective parameters to classify dysphonic and normophonic voices and commonly include auditory-perceptual assessment of voice quality, acoustic and aerodynamic assessment, patient self-ratings, and laryngeal imaging.[10,11] Normative data for acoustic and aerodynamic parameters in male-female differences and for some culturally diverse populations have been established[5,12-14] but a paucity exists in higher level empirical studies for acquired acoustic and aerodynamic measures in non-native English speakers. Auditory-perceptual assessment and acoustic measures can be especially impacted by differences in physiology, accents, cultural, and linguistic differences in speech due to the nature of the assessment. It is important to understand these differences and account for them when performing clinical voice evaluations.

Voice evaluation: auditory-perceptual analysis and accent

Standardized tools for auditory-perceptual evaluation of the voice include the GRBAS Scale created by Hirano, quantifying grade (G), roughness (R), breathiness (B), asthenia (A), and strain (S) on a four-point scale,[15] and the Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V).[16] The language background of an SLP may influence the auditory-perceptual evaluation[17] as evidenced by statistically different perceptual evaluations based on nativity of language via the GRBAS scale for speakers of Japanese, Standard American English (SAE), Canadian English, French, Italian, and Portuguese.[18-20]

The particular pattern of pronunciation within a standard language that may characterize an individual by their sex, social class, age, geographic region, profession, and culture can be defined as a speaker's accent.[21-23] One parameter of auditory-perceptual evaluation is hypernasality, yet linguistically, nasality may be characteristic of a certain accent or provide contrastive phonemic elements within a language. Similarly, differences in other features such as aspiration, glottal fry, intensity level, rate of speech across languages and cultures may suffuse into a speaker's accent when using SAE. While accent is acoustic-phonetic with phonological deviations from the normative language, *dialect* includes grammatical differences in addition to acoustic-phonetic and phonological deviations. Accents and dialects of a language contribute to its unique phonemic expression and the accuracy in delivery of that message within established cultural markers. Unexpected transfer of phonemic attributes from one accent or language to another creates a divergence if enough contrastive elements in temporal, tonal, and dynamic features of the native language occurs.[24] The degree of accentedness in the vocal signal may be influenced by the age of second language learning (L2), length of residence in the L2 country, frequency of continued use of native language, and motivation by the speaker[25] and, conversely may influence the credibility of an individual and perception of professionalism, comprehensibility, and intelligence.[26-30] Given that there are physiological differences between races and ethnicities, as supported by research findings,[7,8,9] and that accents and languages have an effect on voice quality and its perception[21-23,26-30] it is important to establish normative data for persons from different sociolinguistic populations and cultural backgrounds.

Voice evaluation: acoustics

Objective and noninvasive measures of voice evaluation are possible through assessment of acoustic features of the voice. A study by Andrianopoulos et al controlled for dialectal and linguistic variables among four homogeneous groups, native Mandarin-Chinese speakers, native Hindi-Indian speakers, Caucasian-Americans, and African-American speakers of SAE, to measure spectral and fundamental frequency (F_0) in three vowel tasks (sustained /a/, /i/, and /u/).[23] As expected, all female participants across homogeneous groups demonstrated significantly higher fundamental frequencies for the vowel tasks than male participants. Statistically significant differences were found in acoustic and aerodynamic assessments between these sociolinguistic sample populations. Due to the absence of continuous speech from the dependent variables of solely isolated speech tasks, generalization is limited in this study.

Studies have shown significant differences in F_0 and spectra measures between two languages in connected speech stimuli for bilingual proficient speakers.[17,31-33] No omnibus research has been published regarding the evaluation of non-native English speakers in continuous speech stimuli. Additionally, the manipulation of the vocal tract and articulators to produce the phonetic properties unique to a language may vary the voice quality in linguistically diverse speakers.[31,33,34]

Jitter, shimmer, noise-to-harmonic ratio (NHR) were common clinical measures for acoustic voice evaluation and continue to be used even now, but cepstral analysis is now preferred due to its reliability with isolated vowels, connected speech samples and aperiodic samples (no cyclic behavior necessary). Cepstral analysis is also advantageous due to its accuracy in samples with modulation and intermittency, and its management of both noise in consonantal production and high variation in amplitude and pitch concurrently.[35,36] The Analysis of Dysphonia in Speech and Voice (ADSV, Pentax Medical, Montvale, NJ) [35] employs cepstral analyses by using frequency-based acoustic analyses to track the cepstral peak within a cepstrum, computed through a forward Fourier transformation of the logarithmic power spectrum of a recorded sound wave.[35] ADSV is actively used by the practicing clinician and/or researcher as a diagnostic tool, for measurement of progress in treatment, and in the trajectory toward disorder-specific assessment.[37]

Preliminary normative measures for ADSV were solely obtained from North American speakers, and while diverse populations have been studied using ADSV, with one study including native Flemish speakers,[12,38] normative data were not established nor was there disclosure of the inclusion of non-native speakers of English. Commensurate with ADSV, the Multi-Dimensional Voice Program (MDVP, Kay Elemetrics, Montvale, NJ)[39] was also normalized on North American speakers. A cross-sectional study comprised of Saudi male and female speakers revealed 15 out of the 33 MDVP variables and 10 of the 33 MDVP variables, respectively, were significantly different from the established normative values.[40] With research indicating significant differences in acoustic and aerodynamic parameters for nonspeech and speech tasks between culturally and linguistically diverse sample populations, there is a need to ensure that objective measurements accurately interpret the voice quality classification of all individuals.

Purpose

The purpose of this study was to (1) compare objective measures of voice quality assessment of monolingual speakers of SAE with native speakers (L1) of French and Spanish on perturbation, noise, and cepstral/spectral analyses; and (2) compare ratings on auditory-perceptual assessment with acoustic data secondary to degree of accentedness.

METHODS

The University of Houston Institutional Review Board reviewed and approved the prospective cohort study. Three groups of participants completed a one-time assessment of voice quality using auditory-perceptual ratings and acoustic assessment, and an informal assessment of accentedness. Prior to initiating the study procedures, participants completed an informed consent form and a demographic questionnaire, which highlighted inclusionary criteria necessary for participation. Two participants were exempt from the study due to exclusionary responses on the questionnaire related to language acquisition.

Participants

Thirty individuals (10 L1 English, 10 L1 Spanish, and 10 L1 French) participated in the study, 15 male and 15 female participants. They were recruited by native language, sex, age, and proficiency in English. Subjects met the following criteria: (1) healthy individuals on self-report, (2) 18-60 years of age, (3) self-reported normal voice quality with no current or history of a diagnosed voice disorder, neurological, or respiratory disorder, (4) nonsmoking for at least 5 years, (5) monolingual, native speakers of English or native speakers of French or Spanish. Spanish or French speakers will have learned English as a second language after a minimum of 16 years of age, who were born, raised, and educated for at least the first 16 years of his or her life in the same country, (6) elementary and secondary education were provided in the subject's native language and regional dialect. Age and demographic characteristics collected from the language questionnaire are provided in Tables 1 and 2.

Procedures

Subjects were recorded using MDVP on a desktop computer in a double-walled soundproof booth minimizing ambient noise. Participants completed the following tasks: reading of the Rainbow Passage, reading of CAPE-V sentences and sustaining vowel /a/. Participants wore the AKG C520 head-mounted omnidirectional condensor microphone, and mouth-to-microphone distance was set to 10 cm from the lips, at an angle of 45° to 90° away from the front of the mouth, based on ASHA's recommended protocols for instrumental assessment of voice.[11] All recordings were de-identified prior to analysis through a randomization engine. The study personnel (TP and AJ) completed the CAPE-V form to evaluate vocal attributes of all participants, with one author (AJ) blinded to data collection. Voice quality measures for frequency and perturbation measures (jitter, shimmer, NHR) were analyzed using MDVP, and cepstral

TABLE 1.
Mean Age and Age Range for Participants by Native Language

	English		French		Spanish	
	Female (n = 5)	Male (n = 5)	Female (n = 5)	Male (n = 5)	Female (n = 5)	Male (n = 5)
Mean age (years)	28.6	31	37.6	30.2	35	35.2
Range (years)	22-41	24-37	26-69	28-34	29-39	29-46

TABLE 2.
Linguistic Background of Non-native SAE Participants

Demographic Factors	French		Spanish	
	Female n = 5	Male n = 5	Female n = 5	Male n = 5
Mean years (yr; mo) of formal education in L1 (range: yr; mo)	20; 9 (14; 0-34; 0)	21; 2 (17; 0-27; 0)	17; 0 (15; 0-20; 0)	18; 7 (16; 0-24; 0)
Mean years of residency in L1 country	27; 7 (22; 0-35; 0)	23; 7 (20; 0-27; 4)	23; 9 (16; 4-27; 0)	26; 7 (23; 0-28; 0)
Mean years of residency in L2 country	8; 7 (4; 6-25; 0)	6; 10 (3; 4-12; 1)	11; 6 (5; 1-19; 0)	8; 6 (3; 4-23; 10)
Mean age of first exposure to L2	8; 3 (0; 0-10; 0)	10; 6 (9; 0-11; 0)	6; 9 (4; 0-12; 0)	5; 2 (4; 0-6; 0)
Mean age of social proficiency in L2	21; 4 (18; 0-29; 0)	22; 2 (18; 0-29; 0)	20; 0 (16; 0-25; 0)	19; 9 (18; 0-23; 0)
Mean percentage of L2 spoken in daily life	58.2 (20-93)	73.4 (63-86)	66.8 (39-77)	58.8 (24-89)

measures (Cepstral Peak Prominence [CPP] and Cepstral-Spectral Index of Dysphonia [CSID]) using ADSV.

Perceptual measures of accentedness (Figure 1) were obtained informally for each participant on a 100-word sample of a standardized stimulus (The Rainbow Passage). Two graduate students in speech-language pathology, blinded to the purpose of the study and with experience in working with sociolinguistic populations through the accent modification clinic at the University of Houston, rated these recordings. The evaluators participated in an additional 1-hour training to standardize rating parameters and the scale unique to this study. Presence of acoustic-phonetic and phonological deviations from SAE was marked and totaled as a percent out of 100.

Data analyses

Raw data were analyzed for acoustic variables using the *ADSV* and *MDVP* software. Statistical analysis was executed with *SPSS version 24.0* (IBM, Released 2016, Armonk, NY). Statistical analyses included analysis of variance (ANOVA) to examine differences within and between sociolinguistic groups, Fisher's least significance difference post hoc analyses to examine interactions of statistically significant variables, and Pearson's product moment correlation for inter-rater reliability and degree of accentedness with objective variables. Significance was established at $P < 0.05$.

RESULTS

Comparison within groups

Descriptive statistics for all 19 dependent variables of voice quality are reported in Table 3. A univariate ANOVA was performed and group results for the auditory-perceptual, acoustic, and accentedness parameters are reported in Table 4. In age and demographic characteristics, native speakers of Spanish, French, and SAE were commensurate at baseline with no significant differences between these groups.

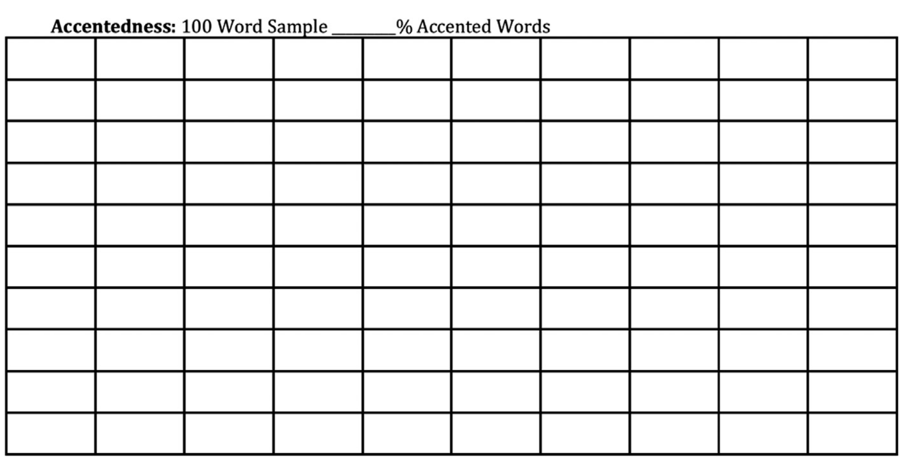

FIGURE 1. Accentedness speech sample.

TABLE 3.
Mean and Standard Deviations for CAPE-V Scores, Perturbation, Noise Measures, and Cepstral/Spectral Measures

1		English		French		Spanish	
		Male (n = 5)	Female (n = 5)	Male (n = 5)	Female (n = 5)	Male (n = 5)	Female (n = 5)
CAPE-V scores	Overall severity	1.60 (1.95)	0.40 (.55)	0.60 (.55)	1.80 (4.03)	1.40 (1.65)	0.80 (1.80)
	Roughness	1.80 (2.17)	0.60 (.89)	1.20 (.84)	1.60 (3.58)	2.20 (.84)	1.00 (2.24)
	Breathiness	0.60 (1.34)	0.00 (0.00)	0.00 (0.00)	0.20 (.45)	1.00 (1.23)	0.20 (.45)
	Strain	0.20 (.45)	0.00 (0.00)	0.00 (0.00)	1.40 (3.13)	0.00 (0.00)	0.40 (.55)
Fundamental frequency (Hz)	F_0	108.94 (10.03)	235.07 (28.67)	119.41 (11.13)	188.10 (42.92)	113.37 (12.42)	204.85 (24.22)
	Speaking F_0	109.60 (8.00)	201.66 (21.48)	120.20 (12.85)	191.15 (27.58)	119.12 (16.48)	193.88 (10.01)
Perturbation (%)	Jitter	0.59 (.21)	1.09 (.67)	0.55 (.28)	1.02 (.59)	1.62 (1.19)	1.36 (.75)
	Shimmer	0.30 (.11)	0.26 (.06)	0.29 (.15)	0.22 (.04)	0.44 (.37)	0.72 (.94)
Noise (dB)	Noise-to-harmonic ratio (NHR)	0.14 (.02)	0.10 (.02)	0.14 (.02)	0.12 (.03)	0.16 (.06)	0.12 (.02)
Cepstral/spectral measures	Sustained vowel CPP	13.67 (1.90)	10.06 (1.30)	13.48 (2.15)	12.32 (2.60)	13.68 (1.61)	12.40 (.80)
	Sustained vowel CSID	-0.68 (10.46)	-2.08 (4.90)	-9.42 (13.19)	-9.31 (9.90)	-5.47 (11.26)	-7.96 (4.95)
	All-voiced sentence CPP	6.98 (.93)	6.55 (1.90)	6.37 (1.14)	4.83 (.83)	5.52 (.29)	5.11 (.61)
	All-voiced sentence CSID	-9.53 (12.08)	-3.61 (11.71)	-6.39 (10.47)	7.97 (10.80)	-1.38 (12.03)	4.09 (5.51)
	Easy onset sentence CPP	8.11 (.62)	8.04 (.64)	8.65 (1.74)	7.42 (1.53)	8.09 (1.53)	7.62 (1.26)
	Easy onset sentence CSID	-7.89 (6.16)	-2.81 (13.20)	-9.14 (6.66)	2.79 (15.34)	-9.94 (12.45)	-8.93 (5.46)
	Hard glottal sentence CPP	5.46 (.56)	4.85 (.72)	5.56 (1.03)	4.64 (0.93)	5.01 (1.14)	4.90 (1.07)
	Hard glottal sentence CSID	1.91 (12.85)	-1.65 (9.25)	-6.70 (12.28)	6.35 (11.94)	-0.24 (18.08)	3.07 (13.50)
	Voiceless plosive sentence CPP	5.27 (1.03)	5.02 (.60)	6.30 (1.50)	4.55 (.77)	5.48 (1.88)	4.68 (.95)
	Voiceless plosive sentence CSID	-0.75 (14.30)	7.24 (8.44)	-9.34 (14.02)	7.52 (14.02)	-1.43 (29.07)	11.82 (8.72)

Comparison between groups

Acoustic analyses

Frequency, jitter, shimmer, NHR, and cepstral/spectral measures (CPP and CSID for sustained vowel and CAPE-V sentences) comprised the acoustic analyses. To answer Hypothesis 1: "do significant differences exist in objective measures of voice quality of monolingual speakers of SAE with native speakers (L1) of French and Spanish on perturbation, noise, and cepstral/spectral analyses," a univariate ANOVA was performed to examine within and between group differences. Significant main effects were found for language for CPP of the all-voiced sentence. There was a significant interaction effect for language for CPP and CSID of the all-voiced sentence, and jitter for the sustained

TABLE 4.
ANOVA Results of CAPE-V Scores and Acoustic Variables by Sex, Native Language, and Interaction of Sex and Language

Measures		Sex			Language			Sex* Language		
		df	F	*P*	df	F	*P*	df	F	*P*
CAPE-V scores	Overall severity	1	0.279	0.602	2	0.093	0.911	2	1.116	0.344
	Roughness	1	0.810	0.377	2	0.097	0.908	2	0.518	0.602
	Breathiness	1	1.946	0.176	2	1.027	0.373	2	1.135	0.338
	Strain	1	1.243	0.276	2	0.602	0.556	2	0.951	0.400
Fundamental frequency (Hertz)	F_0	1	112.913	<0.001*	2	1.454	0.253	2	3.457	0.048*
	Speaking F_0	1	155.049	<0.001*	2	0.008	0.992	2	1.042	0.368
Perturbation (%)	Jitter	1	0.904	0.351	2	3.145	0.061†	2	0.969	0.394
	Shimmer	1	0.111	0.742	2	1.849	0.179	2	0.516	0.604
Noise (dB)	NHR	1	5.431	0.029*	2	1.237	0.308	2	0.289	0.752
Cepstral measures	Vowel CPP	1	9.166	0.006*	2	1.243	0.306	2	1.437	0.257
	Vowel CSID	1	0.129	0.723	2	1.782	0.190	2	0.046	0.955
	All-voiced CPP	1	4.067	0.055†	2	5.099	0.014*	2	0.893	0.423
	All-voiced CSID	1	4.843	0.038*	2	1.715	0.201	2	0.549	0.584
	Easy-onset CPP	1	1.569	0.222	2	0.079	0.924	2	0.515	0.604
	Easy-onset CSID	1	2.401	0.134	2	0.895	0.422	2	0.674	0.519
	Hard-glottal CPP	1	2.606	0.120	2	0.124	0.884	2	0.487	0.620
	Hard-glottal CSID	1	0.777	0.387	2	0.040	0.961	2	0.993	0.385
	Voiceless-plosive CPP	1	4.454	0.045*	2	0.229	0.797	2	0.985	0.388
	Voiceless-plosive CSID	1	4.565	0.043*	2	0.367	0.697	2	0.187	0.830

* Significant ($P < 0.05$).
† Approaching significance.

TABLE 5.
Post Hoc Analyses for Significant Language Interactions by Acoustic Variables

Measures	Language Interactions		P
All-voiced CPP	English	French	0.023*
	English	Spanish	0.006*
	Spanish	French	0.562
Jitter	English	French	0.866
	English	Spanish	0.048*
	Spanish	French	0.034*

* Significant ($P < 0.05$).

TABLE 6.
Percent Scores for Individual Scores on an Informal Rating of Accentedness, on a 100-Word Speech Sample

Participant	Sex	Mean Accentedness	
		French	Spanish
1	F	21.0%	12.5%
2	F	16.0%	28.5%
3	F	29.0%	16.5%
4	F	23.5%	15.0%
5	F	25.0%	27.5%
6	M	7.5%	5.5%
7	M	6.5%	22.5%
8	M	46.5%	21.5%
9	M	42.0%	16.0%
10	M	24.0%	20.0%

TABLE 7.
ANOVA of Accentedness and Language

	Accentedness			Language		
	df	F	P	df	F	P
Jitter	1	0.58	0.457	1	4.702	0.045*

* Significant ($P < 0.05$).

vowel. Post hoc analyses for interaction of language for CPP of the all-voiced sentence, revealed significance between dyads of SAE-Spanish speakers ($P = 0.023$) and SAE-French speakers ($P = 0.006$), shown in Table 5. Native speakers of SAE had significantly higher all-voiced CPP results than non-native SAE speakers. The interaction between French and Spanish speakers for CPP of all-voiced sentence was nonsignificant ($P = 0.562$). While the main effect for jitter by language was approaching significance, post hoc analyses revealed nonsignificance between SAE and French speakers but statistical significance between SAE-Spanish and Spanish-French dyads. Post hoc analyses for interaction of language revealed native speakers of Spanish had significantly higher group mean (1.49%) results in jitter than native French speakers (0.79%) ($P = 0.034$).

A secondary analysis to examine sex differences within and between linguistic groups was performed and a univariate ANOVA revealed significant main effect of sex for fundamental frequency (F_0), speaking F_0, and NHR. Male participants had a significantly lower F_0 (F = 112.913, $P < 0.001$), speaking F_0 (F = 155.049, $P < 0.001$) than female participants as expected, but a higher NHR (F = 5.431, $P = 0.029$). There were no significant differences identified for shimmer. A main effect of sex in cepstral/spectral analyses was observed for CPP of sustained vowel (F = 9.917, $P = 0.006$) and CPP of the voiceless plosive sentence, "Peter will keep at the peak" (F = 4.45, $P = 0.045$). In both of these variables, the group mean of CPP in male subjects was significantly higher than female subjects, indicating increased presence of harmonics in the vocal signal from participating male subjects. There was significant main effect of sex for CSID of the voiceless plosive sentence of the CAPE-V with the group mean of male subjects significantly lower than the female subjects. Main effect of sex was approximating significance for the CPP of the all-voice sentence of the CAPE-V, "We were away a year ago," but there was no significant effect for the remaining stimuli. A sex by language interaction effect was significant for F0 ($P < 0.48$), but further post hoc analyses revealed no significant relationships.

Auditory-perceptual assessment and accentedness

For Hypothesis 2: "do significant differences exist in the auditory perceptual assessment of voice with acoustic data secondary to accentedness?", no statistically significant differences were evidenced for main effect of sex or language for CAPE-V parameters (overall severity, breathiness, roughness, and strain). Each average percentage of accentedness for native speakers of French and Spanish is displayed in Table 6. While controlling for language and accentedness, a univariate ANOVA revealed significant difference, as seen in Table 7 and 8, in jitter ($P = 0.045$) and approaching significance for the all-voiced sentence CPP ($P = 0.056$) and jitter ($P = 0.066$).

Inter-rater reliability was established on 20% of the CAPE-V scores between the two raters (TP and AJ) and, for the accentedness samples (two graduate students) using Pearson's correlation. The Pearson's correlation revealed a strong relationship ($r > 0.95$) indicating good inter-rater reliability for CAPE-V and accentedness. Pearson's correlation revealed no significant relationships between degree of accentedness and objective measures.

DISCUSSION

As the United States continues to diversify, cultural competence within service delivery is necessary, to consider cultural variables and language exposure when providing care to patients and clients.[1] Three homogeneous groups of male and female subjects, a composite of 30 male and female native speakers of SAE, French, and Spanish participated in this preliminary study to assess cultural appropriateness and accuracy in objective measures of voice quality, as well

TABLE 8.
ANOVA of Accentedness, Sex, Language, and the Interaction of Sex and Language

	Accentedness			Sex			Language			Sex* Language		
	df	F	P	df	F	P	df	F	P	df	F	P
Jitter	1	1.010	0.325	1	0.885	0.357	2	3.074	0.066†	2	1.129	0.341
All-voiced CPP	1	1.154	0.294	1	4.047	0.056†	2	1.257	0.303	2	1.059	0.363

* Significant ($P < 0.05$).
† Approaching significance.

as to assess any relationship of auditory-perceptual assessment and acoustic data secondary to degree of accentedness in the non-native SAE speakers as measured by the informal assessment (Figure 1). Hypotheses were (1) no significant differences will exist between monolingual SAE speakers and native speakers of French and Spanish for acoustic measures, and (2) there will be no significant difference between ratings on auditory-perceptual assessment with acoustic data secondary to degree of accentedness.

Of the study's 19 experimental variables related to voice quality and acoustic analyses, significant interaction effects for language were noted for the CPP and CSID for the all-voiced sentence and, jitter for the sustained vowel. Jitter value is considered to be within normal range when it is less than 1%.[13] Statistically significant jitter findings are consistent with cross-sectional studies for culturally and linguistically diverse populations.[23,34] Cepstral measures were obtained in this study since time-based perturbation measures have been shown to have less reliability compared to CPP for identifying dysphonia severity.[12] Both female and male speakers of SAE demonstrated significantly higher group mean CPP for the all-voiced sentence than native French and Spanish speakers. Normative values reported in the application guide for the ADSV[35] indicate the following means (standard deviation [SD]) for the all-voice sentence: for females, mean CPP 7.66 dB (0.95 dB) and mean CSID 6.44 (7.94); for males, mean CPP 8.04 dB (1.33 dB), and mean CSID -4.48 (7.94). The raw data showed that for the all-voiced sentence, one male and one female in each of the native Spanish and French speaking group had CPP values lower than the normative range (mean with 1 SD) reported in the ADSV application guide based on a preliminary sample of North American speakers.[35] For the CSID, a lower value including a negative value is indicative of a good voice quality. When comparing the raw data to the normative CSID values, we found that one male native Spanish speaker and, two female and one male native French speakers had CSID values greater than the normative range, but as indicated in Table 4 all three linguistic groups had large SDs for the CSID. Diagnostically, the all-voiced sentence, "We were away a year ago," assesses the maintenance of "linking" phonetic context together and has previously shown to be the most robust CPP for CAPE-V sentences,[41] with noted sensitivity for aberrations in physiological and acoustical measures in dysphonic voices.[35,42] Structurally, the all-voiced sentence contains one stopped, voiced plosive /g/ in "ago." Work by Watts et al established a normative CPP value of 6.01 dB (SD = 2.44 dB) for the all-voiced sentence CPP with ADSV for North American speakers and a normative value of 9.21 dB (SD = 2.98 dB) for Flemish speakers in their analysis of correlations between algorithms for *ADSV* and *Praat*.[38] It is therefore relevant to establish and utilize normative values for sociolinguistic cultural groups that appropriately represent the patients' background as cross-linguistic variabilities influence objective analyses.

A study by Cantor-Cutiva et al also found significant differences in jitter and harmonics-to-noise ratio measures between male and female monolingual English and bilingual English-Spanish speakers, and differing CPPS values between monolinguals and bilingual speakers using a continuous speech sample stimulus with *Praat* (www.praat.org), but with no statistical significance.[33] As researchers evidenced significant differences in connected speech and F_0 for proficient bilingual speakers,[17,31,32] and timing of execution between languages and voice quality composition (ie, breathiness and nasality),[43,44] cross-linguistic variabilities among sociolinguistic groups appear to impact the presented acoustic signal and must be accounted for during a comprehensive voice evaluation. Results from this study further support the theory that in addition to anatomical differences and supraglottic alterations already defined in sociolinguistic populations,[7,8,9] to effectively communicate in keeping with cultural expectations in each language, a difference may be present in voice production at the glottal level.

Statistically significant differences based on sex were present in cepstral/spectral analyses in CPP measures for sustained vowel and voiceless plosive sentence, and in CSID measures of sustained vowel. In these measures, male subjects had increased presence of harmonics in the voice signal with statistically higher CPP measures and lower CSID measures compared to female participants in both sustained and connected speech stimuli.

Continual work to develop normative data inclusive of diverse, sociolinguistic groups, is warranted, as the literature supports significant differences across populations and in objective analysis. Capturing and analyzing an acoustic signal must therefore acknowledge cross-cultural and cultural influences. A further study of the physiological and functional differences in voice production in native and non-native speakers of SAE will also allow us to better predict how these variations may manifest in the measurement variables.

For Hypothesis 2, there were no significant differences in auditory-perceptual measures of voice in non-native speakers of SAE compared to objective measures. This supported our hypothesis. While audition of an accent in a non-native speaker may influence perceptions by the listener, in this study, the relationship between auditory-perceptual and objective measures for the voice were not impacted by degree of accentedness.

Limitations and future directions

The results of the study were limited by its small sample size. Future directions warrant increased participant numbers per sex by language group to expand upon emerging and significant relationships. Although there were no significant differences in the demographic data between the sociolinguistic groups, heterogeneity in degree of non-native accent in verbal expression of L2 was variable, and each individual has unique experiences in participation with language and culture. Country of origin for native speakers of Spanish, French, and English was specified in the criteria to Mexico, France, and the United States, respectively, but did not specify region. Additionally, other sociolinguistic groups, such as native speakers of a tonal language, were not represented in this study. Exclusionary criteria required a certain level of English proficiency for French and Spanish participants; thereby this is not a complete representation of non-native SAE speakers for objective and perceptual measures. Additionally, only one of two raters were blinded for the auditory perceptual evaluation for the CAPE-V analysis.

Future studies should explore the differences between simultaneous and sequential bilingualism in their impact on auditory perception. While the current study did not show any significant differences between auditory-perceptual and acoustic data, it would be important to assess this relationship for the disordered voice and, for clinicians with lesser experience with sociolinguistic diversity or practicing in regions with limited diversity, as familiarity with a language can affect auditory perception.[26-28,30]

CONCLUSION

In this preliminary study, measures of all-voiced CPP and jitter were statistically significant, demonstrating differences between monolingual and sequential bilingual Spanish and French speakers of English. It is important to recognize that these differences exist between native and non-native speakers of a language while building a normative database for all diagnostic variables in a diverse society. Utilization of all diagnostic tools (ie, perceptual skills and instrumentation) is warranted in creating a complete and accurate diagnostic vocal profile of patients from all cultural and linguistic backgrounds. As delineated by ASHA's Code of Ethics regarding cultural competency, cultural variables, language exposure, and culturally linguistic differences need to be considered in service delivery. For the evaluating clinician, recognizing that nativity of a language may have an effect on the voice signal of the patient, as well as its analyses through objective and subjective measures is increasingly critical.

ACKNOWLEDGMENTS

We would like to thank Dr Laura Cizek and Dr Elizabeth Goodin-Mayeda for their guidance on this study, and Caitlyn Pratt and Jennifer Withner for completing the informal assessment of accentedness.

REFERENCES

1. American Speech-Language-Hearing Association. Issues in ethics: Cultural and linguistic competence. 2017. Available at: www.asha.org/Practice/ethics/Cultural-and-Linguistic-Competence/. Accessed March 20, 2019.
2. United States Census Bureau. Age by language spoken at home by ability to speak English for the population 5 years and older. [Data file]. 2017. Available at: https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=ACS_15_SPT_B16004&prodType=table. Accessed May 30, 2018.
3. Roy N, Merril RN, Gray SD, et al. Voice disorders in the general population: prevalence, risk factors and occupational impact. *Laryngoscope*. 2005;115:1988–1995.
4. Owens R. *Language Development: An Introduction*. 9th ed. New York: Pearson Education; 2015.
5. Awan SN, Mueller P. Speaking fundamental frequency characteristics of white, African American, and Hispanic kindergartners. *J Speech Lang Hearing Res*. 1996;39:573.
6. Berg M, Fuchs M, Wirkner K, et al. The speaking voice in the general population: normative data and associations to sociodemographic and lifestyle factors. *J Voice*. 2017;31:257.e13–257.e24. https://doi.org/10.1016/j.jvoice.2016.06.001.
7. Xue S, Hao J. Normative standards for vocal tract dimensions by race as measured by acoustic pharyngometry. *J Voice*. 2006;20:391–400.
8. Morgan NJ, MacGregor FB, Birchall MA, et al. Racial differences in nasal fossa dimensions determined by acoustic rhinometry. *Rhinology*. 1995;33:224–228.
9. Corey J, Gungor A, Liu X, et al. Normative standards for nasal cross-sectional areas by race as measured by acoustic rhinometry. *Otolaryngol-Head Neck Surg*. 1998;119:389–393.
10. Roy N, Barkmeier-Kraemer J, Eadie T, et al. Evidence-based clinical voice assessment: a systematic review. *Am J Speech-Lang Pathol*. 2013;10:212–226. 1044/1058-0360(2012/12-0014).
11. Patel RR, Awan SN, Barkmeier-Kraemer J, et al. 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*. 2018;27:887–905.
12. Awan SN, Roy N, Cohen S. Exploring the relationship between spectral and cepstral measures of voice and the Voice Handicap Index (VHI). *J Voice*. 2014;28:430–439. https://doi.org/10.1016/j.jvoice.2013.12.008.
13. Baken RJ, Orlikoff RF. *Clinical Measurement of Speech and Voice*. Clifton Park, NY: Cengage Learning; 2010.
14. Zraick RI, Smith-Olinde L, Shotts LL. Adult normative data for the KayPENTAX phonatory aerodynamic system model 6600. *J Voice*. 2012;26:164–176.
15. Hirano M. *Clinical Examination of Voice*. New York: Springer; 1981.
16. Kempster GB, Gerratt BR, Verdolini Abbot K, et al. Consensus auditory-perceptual evaluation of voice: development of a standardized clinical protocol. *Am J Speech Lang Pathol*. 2009;18:124–132.

17. Jarvinen K, Laukkanen A-M, Geneid A. Voice quality in native and foreign languages investigated by inverse filtering and perceptual analyses. *J Voice*. 2017;31:261.e25–261.e31.
18. Yamaguchi H, Shrivastav R, Andrews ML, et al. A comparison of voice quality ratings made by Japanese and American listeners using the GRBAS scale. *Folia Phoniatrica Logopaedica*. 2003;55:147–157.
19. Ghio A, Cantarella G, Weisz F, et al. Is the perception of dysphonia severity language-dependent? A comparison of French and Italian voice assessments. *Logopedics Phoniatrics Vocol*. 2013;40:36–43.
20. Behlau M, Yamaguchi H, Andrews ML. Escala GRBAS em três diferentes culturas. Proceedings of the 9th Brazilian Congress of Speech Language Pathology and Audiology, Guarapari, ES, Brazil, 586; 2001.
21. Giles H. Evaluative reactions to accents. *Educ Rev*. 1970;22:211–227.
22. Wells J. *Accents of English*. Cambridge: Cambridge University Press; 2012.
23. Andrianopoulos MV, Darrow KN, Chen J. Multimodal standardization of voice among four multicultural populations: fundamental frequency and spectral characteristics. *J Voice*. 2001;15:194–219.
24. Trofimovich P, Baker W. Learning second language suprasegmentals: effect of L2 experience on prosody and fluency characteristics of L2 speech. *Stud Second Lang Acquisition*. 2006;28:1–30. https://doi.org/10.1017/s0272263106060013.
25. Piske T, MacKay I, Flege J. Factors affecting degree of foreign accent in an L2: a review. *J Phonetics*. 2001;29:191–215.
26. Derwing TM, Munro MJ. Accent, intelligibility, and comprehensibility: evidence from four L1s. *Stud Second Lang Acquisition*. 1997;19:1–16.
27. Munro MJ, Derwing TM. Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Lang Learn*. 1995;45:73–97.
28. Fuse A, Navichkova Y, Alloggio K. Perception of intelligibility and qualities of non-native accented speakers. *J Commun Disord*. 2018;71:37–51.
29. Carlson H, McHenry MA. Effect of accent and dialect on employability. *J Employment Counsel*. 2006;43:70–83.
30. Flege JE. Factors affecting degree of perceived foreign accent in English sentences. *J Acoust Soc Am*. 1988;84:70–79.
31. Ng M, Chen Y, Chan E. Differences in vocal characteristics between Cantonese and English produced by proficient Cantonese-English bilingual speakers—a long-term average spectral analysis. *J Voice*. 2012;26:171–176.
32. Bahmanbiglu SA, Mojiri F, Abnavi F. The impact of language on voice: an LTAS study. *J Voice*. 2017;31:249.e9–249.e12.
33. Cantor-Cutiva LC, Bottalico P, Nudelman C, et al. Do voice acoustic parameters differ between bilingual English-Spanish speakers and monolingual English speakers during English production? *J Voice*. 2019. https://doi.org/10.1016/j.jvoice.2019.08.009. In press.
34. Esling J, Wong R. Voice quality settings and the teaching of pronunciation. *TESOL Quart*. 1983;17:89–95. https://doi.org/10.2307/3586426.
35. Awan SN. *Analysis of Dysphonia in Speech and Voice (ADSV): An Application Guide*. Montvale, NJ: Pentax Medical; 2011.
36. Dejonckere P, Bradley Patrick, Clemente Pais, et al. A basic protocol for functional assessment of voice pathology, especially for investigating the efficacy of (phonosurgical) treatments and evaluating new assessment techniques − guideline elaborated by the Committee on Phoniatrics of the European Laryngological Society (ELS). *Eur Arch Otorhinolaryngol*. 2001;258:77–82.
37. Gillespie A, Dastolfo C, Magid N, et al. Acoustic analysis of four common voice diagnoses: moving toward disorder-specific assessment. *J Voice*. 2014;28:582–588.
38. Watts C, Awan SN, Maryn Y. A comparison of Cepstral Peak Prominence measures from two acoustic analysis programs. *J Voice*. 2017;31:387.e1–387.e10.
39. Pentax Medical. Software Instruction Manual, Multi-Dimensional Voice Program (MDVP) Model 5105 (2008). Montvale, NJ: Pentax Medical.
40. Malki KH, Al-Habib SF, Hagr AA, et al. Acoustic analysis of normal Saudi adult voices. *Saudi Med J*. 2009;30:1081–1086.
41. Awan SR, Roy N, Jetté M, et al. Quantifying dysphonia severity using a spectral/cepstral-based acoustic index: comparisons with auditory-perceptual judgements from the CAPE-V. *Clin Linguis Phon*. 2010;24:742–758. https://doi.org/10.3109/02699206.2010.492446.
42. Watts CR. The effect of CAPE-V sentences on cepstral/spectral measures in dysphonic speakers. *Folia Poniatr Logop*. 2015;67:15–20.
43. Styler W. On the acoustical features of vowel nasality in English and French. *J Acoust Soc Am*. 2017;142:2469–2482. https://doi.org/10.1121/1.5008854.
44. Garellek M, Ritchart A, Kuang J. Breathy voice during nasality: a cross-linguistic study. *J Phonetics*. 2018;59:110–121.