Establishing normative voice characteristics of younger and older adults

Jessica Banh[1], Konstantin Naumenko[1], Huiwen Goy[1],
Pascal van Lieshout[1,2,3,4], David N. Fernandes[2,3], Kathy Pichora-Fuller[1,3]
[1]Dept. of Psychology, University of Toronto, Mississauga, 3359 Mississauga Rd North, Mississauga, ON, Canada L5L 1C6
[2] Dept. of Speech-Language Pathology, University of Toronto, 500 University Ave, Toronto, ON, Canada M5G 1V7
[3]Toronto Rehabilitation Institute, 550 University Ave, Toronto, ON, Canada M5G 2A2
[4]Inst. of Biomaterials and Biomedical Engineering, University of Toronto, 164 College St, Toronto, ON, Canada M5S 3G9

1. INTRODUCTION

The use of our voice is an important element of daily social interaction. The mechanism of voicing is complex and there are many disorders and hazards that can disrupt it, such as smoking and degenerative neurological diseases. The loss of voice quality and strength is often seen as a result of physical deterioration, but may in fact be a biological precursor for various pathologies [1]. In order to use voice characteristics for diagnostic purposes clinicians need reliable reference values from normal populations to determine the nature and level of impairment.

An important step towards establishing such reference values is to conduct normative studies on a relatively large sample of healthy individuals across age and gender groups. This project is an important contribution towards this aim. It used a new and unique program for recording voice samples called *SpeechClinic*, which was created by Fernandes and Van Lieshout as a new clinical tool to perform multidimensional voice analysis [2]. When a voice sample is recorded, SpeechClinic displays calculated values, such as average pitch, length of phonation and sound pressure level for each task, in comparison with values represented in a unique and flexible database based on a large selection of published references. The algorithms used by SpeechClinic in its analyses are well-established, and the parameters measured are therefore directly comparable with most values in the literature. Finally, SpeechClinic allows the presentation of this information over time for a given patient, which is useful in monitoring treatment progress.

There are two main reasons for conducting this study. First, normative acoustic data do not exist for all voice characteristics measured by SpeechClinic, and the ones that do exist are often based on small sample sizes. Without good normative data, judgments based on these measurements may be inaccurate. Second, computer analysis of voice samples is gaining interest in clinics, but unlike SpeechClinic, most of the existing software lacks an intuitive interface, a well-documented reference database that can be easily extended, or facilities to easily document progress and print reports. This study was able to test SpeechClinic extensively on these features. The primary purpose of this study was to obtain a large number of acoustic voice measures for healthy younger and older adults to provide critical information on the typical distribution of these voice characteristics. In addition, the study allowed us to analyze age-related and gender differences more specifically. The knowledge of age-dependent changes in vocal properties is important in establishing norms because it separates the "normal" from "abnormal" population without confounding typical age-related changes with declines due to medical conditions. After a set of norms is obtained for a particular population, the data from an individual belonging to the same group can be compared to the distribution characteristics obtained from that population. Future studies will focus on sampling other populations to broaden the scope of the database. The results presented here are preliminary findings on a part of the targeted population sample.

2. METHOD

2.1 Participants

Data for a total of 169 people are presented here. The younger group contains 111 first year psychology students, including 55 females (median = 18.6 yrs, range 18-23) and 56 males (median = 18.8 yrs, range 18-27). The older group contains 58 participants recruited from a pre-existing volunteer database, including 47 females (median = 70.7 yrs, range 65-81) and 11 males (median = 73.4 yrs, range 66-78). Participants were included if they learned English before the age of 5 years in an English-speaking country, had pure-tone audiometric thresholds $\leq$25 dB HL from 0.25 to 3 kHz in both ears and no voice-altering conditions (e.g., colds, allergies) at the time of the experiment, and were free of self-reported speech disorders and voice pathologies.

2.2 Apparatus

The equipment used to record voice samples included a SHURE Beta 54 wrap-around headset microphone, which was connected to an M-Audio Mobile Pre USB pre-amplifier. The signal was output to a Mac OS X laptop computer running SpeechClinic. A Korg CA-30 chromatic tuner and a Source brand sound level meter were used for calibration. Voice samples were recorded in an Industrial Acoustics Company double-walled sound-attenuating booth. A GSI-61 clinical audiometer was used to measure the hearing thresholds of younger participants, while older participants already had data on hearing thresholds from earlier recent screenings.

2.3 Procedure

Each participant was tested on six tasks which were performed three times in a row, with task order counterbalanced. Participants phonated the vowel [a] for maximum phonation time, at maximum pitch level, and at the lowest sound pressure level (SPL), and they also read a short passage. These four tasks were preceded and followed by phonating the vowel [a] at habitual frequency and intensity for 8 seconds.

3. RESULTS

Table 1 summarizes the means and standard deviations (SD) of acoustic measures for all 169 subjects, separated by age and gender. Table 2 shows the results from a Mann-Whitney U test for gender and age differences. Three measures of jitter (local, RAP, PPQ), four measures of shimmer (local, APQ3, APQ5, and APQ11), harmonics-to-noise ratio (HNR)[1], habitual SPL, and average fundamental frequency (F_0) were taken from the first habitual (baseline) task, based on 4-sec intervals determined automatically by the program.

Table 1. Means of acoustic measures for two age and gender groups (with standard deviations in parentheses).

	GROUP			
MEASURE	**Younger Female**	**Younger Male**	**Older Female**	**Older Male**
Habitual F_0 (Hz)	254 (30)	130 (21)	215 (41)	119 (22)
Maximum F_0 (Hz)	472 (117)	411 (118)	514 (114)	411 (105)
Max / Habitual F_0	1.87 (0.47)	3.20 (0.96)	2.43 (0.55)	3.50 (0.79)
Jitter local (%)	0.38 (0.29)	0.37 (0.12)	0.45 (0.42)	0.48 (0.19)
Jitter RAP (%)	0.22 (0.19)	0.21 (0.08)	0.26 (0.25)	0.25 (0.12)
Jitter PPQ (%)	0.22 (0.14)	0.21 (0.06)	0.24 (0.23)	0.28 (0.11)
MPT (sec)	14.0 (4.3)	17.8 (5.8)	17.7 (5.8)	17.8 (5.1)
Habitual SPL (dB)	76.2 (3.5)	74.5 (3.2)	74.8 (4.5)	72.1 (4.7)
Minimum SPL (dB)	60.4 (5.8)	55.3 (5.3)	55.3 (9.7)	56.8 (5.8)
(Hab - Min) / Hab SPL	0.21 (0.06)	0.26 (0.06)	0.26 (0.13)	0.21 (0.06)
Shimmer local (%)	2.2 (0.8)	2.6 (1.0)	2.6 (1.5)	4.4 (2.6)
Shimmer APQ3 (%)	1.2 (0.5)	1.4 (0.6)	1.5 (0.8)	2.4 (1.5)
Shimmer APQ5 (%)	1.3 (0.5)	1.6 (0.6)	1.5 (0.7)	2.7 (1.6)
Shimmer APQ11 (%)	1.5 (0.5)	2.1 (0.7)	1.7 (0.7)	3.4 (1.9)
HNR (dB)	24.5 (3.0)	22.9 (2.8)	23.8 (3.7)	21.2 (4.6)

Table 2. *P*-values from comparisons of means using the Mann-Whitney U test. Bolded values highlight significant age- or gender-related differences ($p < 0.01$).

	GROUPS COMPARED			
MEASURE	**Younger F vs. M**	**Older F vs. M**	**Female Y vs. O**	**Male Y vs. O**
Habitual F_0 (Hz)	-	-	**< 0.001**	0.112
Maximum F_0 (Hz)	-	-	0.057	0.960
Max F_0 / Habitual F_0	**<0.001**	**<0.001**	**<0.001**	0.335
Jitter local (%)	0.058	0.129	0.949	0.081
Jitter RAP (%)	0.380	0.326	0.997	0.264
Jitter PPQ (%)	0.072	0.064	0.949	0.044
MPT (sec)	**<0.001**	0.945	**<0.001**	0.820
Habitual SPL (dB)	**0.010**	0.094	0.111	0.048
Minimum SPL (dB)	**<0.001**	0.960	**0.001**	0.672
SPL decrease ratio	**<0.001**	0.208	0.011	0.029
Shimmer local (%)	0.014	0.019	0.278	0.012
Shimmer APQ3 (%)	0.121	0.030	0.297	0.014
Shimmer APQ5 (%)	0.012	**0.004**	0.479	**0.008**
Shimmer APQ11 (%)	**<0.001**	**0.003**	0.146	0.034
HNR (dB)	**0.001**	0.036	0.446	0.136

3.1 Gender differences

On average, younger and older adult males showed a greater increase in pitch in the maximum pitch task than their female counterparts. Younger males also showed a longer MPT than younger females, and younger males were able to soften their voice more than younger females in the minimum SPL task. Adult males of both age groups showed higher shimmer values than adult females.

3.2 Age differences

The data revealed lower pitch values for older females compared to their younger counterparts, but there was no difference between younger and older males. Older females had a longer MPT than younger females, but there were no differences between younger and older males. Older males exhibited higher mean shimmer values than younger males. In the current study, jitter was generally stable and did not change with age or gender.

3.3 In comparison to the literature

Our MPT values for young adults were lower than most mean values reported from other studies [4], while we had higher values for HNR [5] and female F_0 [6]. Our jitter values were comparable to those in other studies [7].

4. DISCUSSION

Our preliminary collection of voice samples creates a first impression for several voice characteristics that are currently not represented in the literature. Furthermore, our data show age and gender differences that warrant further investigation. Although these are preliminary data, the ultimate sample will be large enough to derive important distributional characteristics for a large number of voice parameters for healthy individuals. These results will be useful for future studies using SpeechClinic (e.g., for evaluating the impact of smoking or vocal training). We also plan to repeat this study with background noise to mimic more realistic situations found in clinical settings and increase the ecologically validity of the reference values. Finally, the study has shown the usefulness and robustness of SpeechClinic as a clinical and research tool.

REFERENCES

[1] Oguz, H, Demirci, M, Safak, MA, Arslan, N, Islam, A, & Kargin, S (2007). Effects of unilateral vocal cord paralysis on paralysis on objective voice measures obtained by Praat. *Eur Arch Otorhinolaryngol*, 264:257–261.
[2] Fernandes, D & Van Lieshout, P (2008). Assessment software for voice clinicians. American Speech-Language Hearing Association Convention, November 19-22, Chicago, IL. (Poster Presentation)
[3] Boersma, P. (1993). Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. IFA Proceedings 17, 97-110.
[4] Wuyts, FL, De Bodt, MS, Molenberghs, G, Remacle, M, et al. (2000). The Dysphonia Severity Index: An objective measure of vocal quality based on a multiparameter approach. *J Speech Lang Hear Res, 43*(3), 796-809.
[5] Wolfe, V, Fitch, J, & Cornell, R (1995). Acoustic prediction of severity in commonly occurring voice problems. *J Speech Hear Res, 38*(2), 273-279.
[6] Ferrand, CT (2006). Relationship between masking levels and phonatory stability in normal-speaking women. *J Voice 20*(2), 223-228.
[7] Gelfer, MP and Fendel, DM (1995). Comparisons of jitter, shimmer, and signal-to-noise ratio from directly digitized versus taped voice samples. *J Voice*, 9(4), 378-382.

[1] The measure of HNR in SpeechClinic is defined as the ratio of the amount of energy in the harmonic component of the signal to the amount of energy in the noise component as determined by autocorrelation [3]. However, this quantity is not comparable to the quantity reported by MDVP (KayPENTAX) as “HNR”.