Corpus of deaf speech for acoustic and speech production research

Lisa Lucks Mendel, Sungmin Lee, Monique Pousson, Chhayakanta Patro, Skylar McSorley, Bonny Banerjee, Shamima Najnin, and Masoumeh Heidari Kapourchali

Citation: The Journal of the Acoustical Society of America **142**, EL102 (2017); doi: 10.1121/1.4994288

View online: http://dx.doi.org/10.1121/1.4994288

View Table of Contents: http://asa.scitation.org/toc/jas/142/1

Published by the Acoustical Society of America

Corpus of deaf speech for acoustic and speech production research

Lisa Lucks Mendel,[1,a)] Sungmin Lee,[1] Monique Pousson,[1] Chhayakanta Patro,[2] Skylar McSorley,[1] Bonny Banerjee,[3,b)] Shamima Najnin,[3,b)] and Masoumeh Heidari Kapourchali[3,b)]

[1]*School of Communication Sciences and Disorders, University of Memphis, Memphis, Tennessee 38152, USA*

[2]*Heuser Hearing Institute, Louisville, Kentucky 40203, USA*

[3]*Institute for Intelligent Systems, University of Memphis, Memphis, Tennessee 38152, USA*

lmendel@memphis.edu, slee18@memphis.edu, mpousson@memphis.edu, chhayakantpatro@gmail.com, smmcsrly@memphis.edu, bbnerjee@memphis.edu, snajnin@memphis.edu, mhdrkprc@memphis.edu

Abstract: A corpus of recordings of deaf speech is introduced. Adults who were pre- or post-lingually deafened as well as those with normal hearing read standardized speech passages totaling 11 h of .wav recordings. Preliminary acoustic analyses are included to provide a glimpse of the kinds of analyses that can be conducted with this corpus of recordings. Long term average speech spectra as well as spectral moment analyses provide considerable insight into differences observed in the speech of talkers judged to have low, medium, or high speech intelligibility.

[AL]

Date Received: May 9, 2017 **Date Accepted:** June 28, 2017

1. Introduction

The speech of individuals who are deaf is unique in many ways and has been characterized by several notable features that are distinct to this population (e.g., Osberger and McGarr, 1982; Hedrick *et al.*, 2011). Individuals with hearing impairment exhibit numerous error patterns including omission, substitution, and place of articulation errors, and detailed analysis of these errors can provide important information regarding these individuals' hearing deficiencies and subsequent production deficits (e.g., Najnin *et al.*, 2016; Banerjee *et al.*, 2016).

There is a paucity of recordings of deaf speech available for such acoustic analysis and subsequent applied research. Enhancing our understanding of the unique acoustic features of deaf speech can be very useful for finding ways to improve such individuals' speech perception. This information can be used to help refine how hearing instruments are adjusted to enhance important missing speech information. Successful tuning of such amplification devices can ultimately provide clearer speech input, improved speech understanding, and ultimately better-quality speech production for individuals who are deaf.

A corpus of recordings of speech in a General American dialect is presented here with the aim of informing the larger community of the availability of such soundtracks. Individuals with normal hearing as well as those with severe-to-profound hearing loss read several passages of speech to produce these recordings. A description of the participants, the recordings, and preliminary analyses follows.

2. Data acquisition

2.1 Participants

The participant pool included 40 adults with a General American Dialect divided into two groups. The hearing-impaired (HI) group consisted of 30 adults (22 females; 8 males) between the ages of 16 and 77 years [mean (M): 50, standard deviation (SD): 17]. The HI participants had at least a severe sensorineural hearing loss bilaterally with a minimum pure-tone average of 70 dB hearing level (HL) in their better hearing ear. Some of the HI participants considered themselves part of Deaf culture (N = 12) while others used oral speech and language and were part of the hearing community. These HI participants had poor speech production capabilities as evidenced by poor

[a)]Author to whom correspondence should be addressed.

[b)]Also at Department of Electrical and Computer Engineering, University of Memphis, Memphis, Tennessee 38152, USA

performance on the Computerized Articulation and Phonology Evaluation system (CAPES; Masterson and Bernhardt, 2001) and were in reportedly good physical health with no physical, mental, cognitive or emotional limitations. Table 1 provides a summary of the HI participants' demographic information.

The normal hearing (NH) group comprised 10 adults (6 females; 4 males) aged 15 to 51 years (M: 27, SD: 11). Individuals in the NH group had hearing within a normal limit with a pure-tone average better than 20 dB HL. The NH participants were considered to have speech production skills within a normal range based on their CAPES scores. No known neurological deficiencies were reported for any of the participants. Prior to testing, each participant reviewed and signed the informed consent and completed a hearing history questionnaire. The consent form and the questionnaire were approved by the Institutional Review Board at the University of Memphis. Figure 1 displays a composite audiogram for all participants.

2.2 Stimuli

All participants read five standardized passages that contained words and sentences designed to provide a representative sample of sounds as they occur in the English language. The passages included, "The North Wind and the Sun" (NWS; Aesop, 1999);

Table 1. Demographic information for the hearing impaired participants.

Participant[a]	Age (years)	Gender	Speech Intelligibility Classification	Onset of hearing loss	Age of first amplification use (years)		Current Type of Amplification		Communication Mode[b]
					Right	Left	Right	Left	
1	37	Male	High	Postlingual	5	5	CI[c]	CI	Oral
2	53	Female	Low	Prelingual	42	52	CI	CI	Oral and Sign
3	42	Female	Low	Postlingual	5	5	HA[d]	HA	Oral and Sign
4	38	Female	High	Postlingual	28	NA	CI	NA[e]	Oral
5	16	Female	High	Prelingual	.5	2	HA	CI	Oral
6	77	Male	Medium	Postlingual	18	73	HA	CI	Oral and Sign
7	66	Male	Medium	Postlingual	40	40	HA	HA	Oral and Sign
8	47	Female	High	Postlingual	34	34	HA	HA	Oral and Sign
9	62	Female	High	Postlingual	38	38	CI	HA	Oral and Sign
10	60	Male	High	Postlingual	52	52	HA	CI	Oral and Sign
12	57	Female	Medium	Postlingual	3	58	HA	CI	Oral
13	45	Male	Low	Prelingual	NA	NA	NA	NA	Oral and Sign
14	67	Female	Low	Prelingual	NA	NA	NA	NA	Oral and Sign
15	66	Female	Medium	Postlingual	6	NA	HA	NA	Oral and Sign
16	50	Female	Low	Prelingual	NA	NA	NA	NA	Oral and Sign
17	75	Female	Low	Postlingual	NA	NA	NA	NA	Oral and Sign
18	44	Female	Low	Postlingual	12	12	HA	HA	Sign only
19	48	Female	Low	Prelingual	NA	NA	NA	NA	Sign only
20	47	Female	Low	Prelingual	NA	NA	NA	NA	Sign only
21	47	Female	Low	Prelingual	NA	NA	NA	NA	Sign only
22	58	Female	Low	Postlingual	NA	3	NA	HA	Sign only
23	30	Male	Medium	Postlingual	NA	NA	NA	NA	Sign only
24	33	Male	Low	Postlingual	NA	3	NA	CI	Sign only
25	68	Female	Medium	Postlingual	30	NA	HA	NA	Sign only
26	54	Male	Medium	Prelingual	NA	NA	NA	NA	Sign only
27	51	Female	Medium	Postlingual	7	7	HA	HA	Oral and Sign
28	54	Female	Low	Prelingual	5	7	NA	NA	Sign only
29	55	Female	Medium	Prelingual	12	NA	HA	NA	Sign only
30	71	Female	Medium	Prelingual	NA	55	NA	HA	Oral and Sign
31	49	Female	Medium	Prelingual	2	NA	NA	NA	Sign only

[a]All participants who had either a hearing aid or cochlear implant received (re)habilitation therapy. Participant 11 is not included because he was not an adult.

[b]Oral indicates that the participant used oral speech and language. Sign indicates that the participant used sign language.

[c]Cochlear implant (CI).

[d]Hearing aid (HA).

[e]Not Applicable (NA).

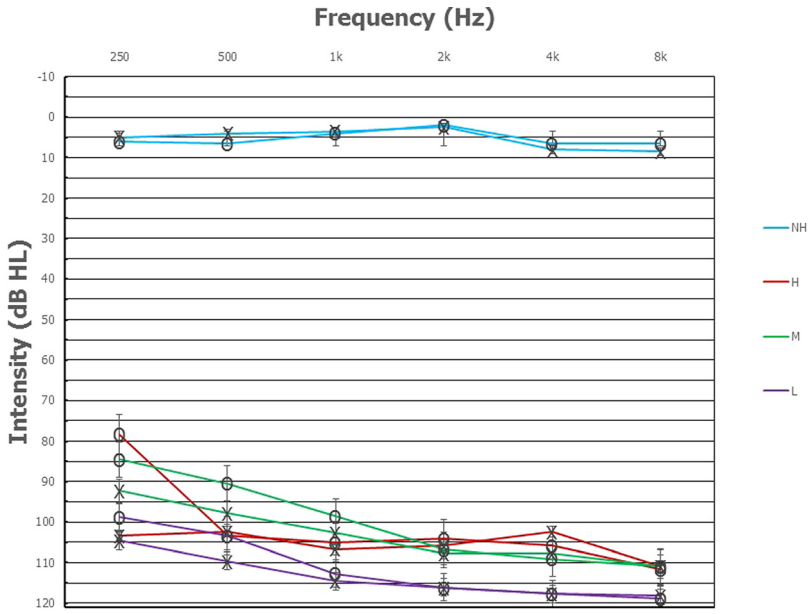

Fig. 1. (Color online) Composite audiogram for the NH and HI groups. "X" refers to left ear thresholds and "0" refers to right ear thresholds.

"The Grandfather Passage" (GP; Van Riper, 1963); "The Rainbow Passage" (RP; Fairbanks, 1960); "Arthur the Rat" (AR; Cassidy, 1985); and "Comma Gets a Cure" (CGC; Honorof *et al.*, 2000). The passages ranged in length from 119 to 593 words each. Recordings were produced while the participants read each passage aloud.

2.3 Instrumentation

The hearing testing and speech production recordings were conducted in a double-walled sound-treated booth meeting ANSI Standard S3.1–1999 (ANSI, 2013) for maximum permissible ambient noise levels for audiometric test rooms. Hearing thresholds were measured using a GSI 61 audiometer and supra-aural TDH-50 headphones. Middle-ear function was assessed using a Maico MI 34 tympanometer. The participants' speech was recorded in both .wav and .mp3 formats. The participants were seated about 10 in. from a Shure SM93 prologue dynamic microphone on a stand, and their speech was recorded in .wav format at a sampling rate of 48 kHz. At the same time, .mp3 recordings were produced using a lapel condenser microphone clipped to the participants' clothing approximately 3 in. from the mouth. Each microphone was connected to separate Marantz Model PMD660 portable solid state recorders to produce the .wav and .mp3 recordings, respectively.

2.4 Procedure

Hearing sensitivity was determined for all participants in each group by measuring pure-tone thresholds at the octave frequencies from 500 to 8000 Hz. After measuring their hearing, the participants were seated in the center of the sound treated booth and instructed to read the passages clearly and naturally. Each passage took approximately 20 min for the participants to complete. Breaks were given at the end of a passage as needed.

3. Description of the corpus

The corpus of recordings described here includes 11 h of speech recordings from the HI and NH groups. The HI participants did not always read fluently and took more pauses than the NH participants, thus they required more time to read the passages (~19 min) compared to the NH participants (~9 min). A total of 660 min of speech was recorded. The recordings were edited making sure all pauses between words and/or non-speech utterances (coughing, throat clearing, etc.) were removed so that the acoustic analyses could be completed on only the speech utterances.

The recordings produced by the HI speakers were classified according to the degree of intelligibility of their speech. Two experienced listeners subjectively rated each participant's speech on a scale from 0 to 7, with 0 referring to completely unintelligible and 7 referring to extremely intelligible. On the basis of this analysis, participants' recordings were classified into three categories: high (extremely intelligible;

rating of 6 or 7), medium (some words were understandable and some were not; rating of 4 or 5), and low (unintelligible; rating of 1, 2, or 3).

The high intelligibility (H) group (N = 6) included individuals who were postlingually deafened and used oral speech and language as their primary mode of communication. Six of the individuals in the medium intelligibility (M) group (N = 11) used oral speech and language (5 postlingual, 1 prelingual), while 5 (3 postlingual, 2 prelingual) primarily used sign language to communicate. All of the participants in the low intelligibility (L) group (N = 13) were nonverbal, and all but one were prelingually deafened.

4. Acoustic Analyses

4.1 Long term average speech spectra (LTASS)

Each participant's recordings of the five passages were analyzed acoustically to display the frequency spectra of their speech production output. Prior to conducting the spectral analyses, each of the recordings was edited using Adobe Audition 3.0 software to remove all silences between utterances. In addition, all extraneous noises, such as loud breaths, throat clearing, and coughing, were removed. After this initial editing, the amplitude level of each participant's vocal output was equalized to a total RMS power of −23 dB. LTASS were then created for the four intelligibility groups (NH, H, M, and L) and are shown in Fig. 2. The average amplitude levels in dB of the five passages as a function of $\frac{1}{3}$ octave frequencies were derived with a 65 536 point Fast-Fourier-Transform (FFT) using a Hamming window. The dB values are relative to the maximum possible level of 0 dB FS (full-scale). Figure 2 indicates that the LTASS for each group was relatively similar.

4.2 Spectral moment analysis

Spoken passages from each speaker were converted into a spectrum (LTASS) displaying the overall speech sample using an FFT. Four spectral moments were derived for the entire passage from each LTASS using Praat (Boersma and Weenink, 2013) which quantitatively investigated the spectral properties of each spoken passage across the four intelligibility groups. *Spectral mean* identified the center frequency where the spectral energy in the signal was greatest. *SD* described the variance among the frequencies represented in the spectrum. *Skewness* provided the degree of spectral tilt in the spectrum, and *kurtosis* indicated the shape of the peak of the distribution.

Figure 3 shows the mean spectral moments for the 5 passages combined for each of the 4 intelligibility groups. Figures 3(A), 3(B), 3(C), and 3(D) represent spectral mean, SD, skewness and kurtosis, respectively. A one-way analysis of variance (ANOVA) was conducted to examine whether there were statistical differences in each spectral moment component across the groups categorized by speech intelligibility for the combined passages. Because of the unequal sample size and corresponding

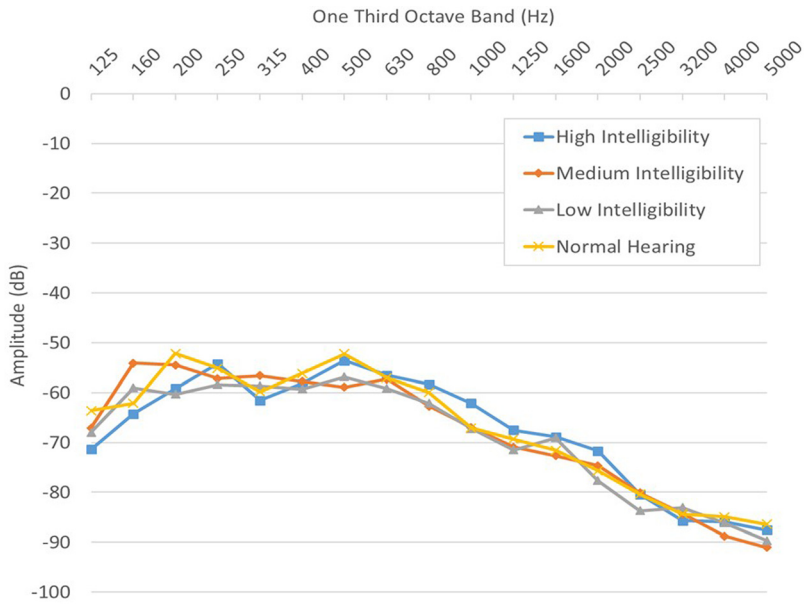

Fig. 2. (Color online) LTASS for the NH and HI groups (H, M, and L) for all five passages combined.

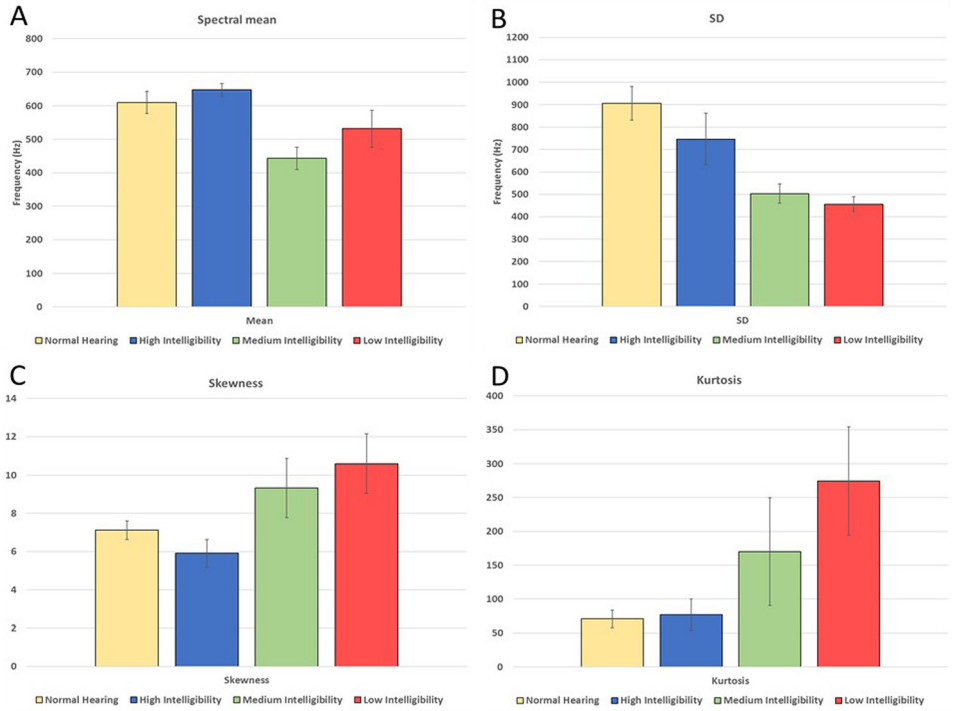

Fig. 3. (Color online) Four spectral moments for the NH and HI groups (H, M, and L) for all five passages combined. (A) Average spectral mean; (B) SD; (C) skewness; and (D) kurtosis. Error bars represent ±1 standard deviation.

violation of homogeneity of variances, a Brown-Forsythe test was used. A Games-Howell statistic was used for *post hoc* analysis with an alpha level of $p = 0.05$.

A significant main effect was found for all four spectral moments: Mean [$F(3,36) = 3.711$, $p = 0.02$], SD [$F(3,36) = 10.415$, $p = 0.00$], skewness [$F(3,36) = 3.149$, $p = 0.037$], and kurtosis [$F(3,36) = 3.865$, $p = 0.017$]. The spectral means for the NH and H intelligibility groups were significantly higher than the M group. The SD for the NH group was significantly higher than that of the M and L groups. Skewness for the L group was significantly greater than for the H group. Last, the L intelligibility group showed higher kurtosis than the NH or H intelligibility group. Similar statistical relationships were found when each passage was analyzed separately across the four spectral moments.

These findings showed systematic patterns associated with the different speech intelligibility groups. Specifically, the second moment, SD, increased when the intelligibility of speech increased [Fig. 3(B)]. This suggests the speech produced by individuals who are more intelligible include a broader range of frequencies than the speech produced by those who are less intelligible. In addition, the fourth moment, kurtosis, decreased with increases in intelligibility [Fig. 3(D)] which also suggests that speech that is more intelligible has a broad frequency distribution. The patterns observed among the four intelligibility groups for both SD and kurtosis are consistent in that both suggest that speech that is more intelligible includes a broad range of frequencies compared to less intelligible speech.

The other two moments, spectral mean and skewness, also showed meaningful patterns as a function of intelligibility despite more variability across the groups. Although the spectral mean for the H group was slightly greater than that for the NH group, the spectral mean for both the NH and H groups was considerably higher than that of the two poorer groups (M and L). Thus, individuals with more intelligible speech appear to produce more energy in the high-frequency region. In addition, despite slightly higher skewness for the NH group compared to the H group, overall outcomes suggest that the skewness (location of spectral energy) for the two higher intelligibility groups (NH and H) was considerably lower than the poorer intelligibility groups (M and L). This would imply that the frequency distribution for the high intelligibility groups had a tail toward the left also resulting in greater energy in the high-frequency region. These findings for spectral mean and skewness are consistent with each other suggesting that those with highly intelligible speech produce more energy in the high-frequency regions which contributes to their speech intelligibility.

Our findings suggest that spectral moments can be used to describe deaf speech, however caution should be used in assuming a causal relationship between these acoustic coefficients and one's relative speech intelligibility. There are other factors besides the acoustics of speech that contribute to overall speech production ability. Participants in this study, for example, varied in the amount of intervention they received, and our data suggest that those who wore amplification (hearing aids or cochlear implants) and/or received speech and language therapy as a child had higher speech intelligibility. Future analyses are planned to examine the role of these intervention factors and how they relate to measured acoustic coefficients.

5. Conclusion

The intention of this paper was to introduce the availability of a corpus of recordings of deaf speech to provide a glimpse into the kinds of acoustic analyses that can be conducted with such recordings. Although the LTASS provides an overview of the acoustic energy in the utterances as a function of frequency, the use of spectral moment analysis provides considerably more insight into differences among the different groups of talkers. Individuals with better speech production abilities who are perceived as being highly intelligible (NH and H groups) produce more energy in the high-frequency regions than those whose speech is less intelligible as evidenced by high spectral means and low skewness. In addition, those with greater intelligibility produce speech covering a larger number of frequencies as evidenced by large SDs and low kurtosis compared to less intelligible deaf speakers. These recordings are available from the Speech Perception Assessment Laboratory http://www.memphis.edu/spal/index.php.

Acknowledgments

Appreciation is expressed to Eugene H. Buder for his expertise in the spectral moments analysis. The authors acknowledge support from NSF grant IIS-1231620 and the Herff College of Engineering.

References and links

Aesop (**1999**). "The north wind and the sun," in *Handbook of the International Phonetic Association* (Cambridge University Press, Cambridge, UK).

American National Standards Institute (ANSI) (**2013**). *S3.1-1999, Maximum Permissible Ambient Noise Levels for Audiometric Test Rooms* (American National Standards Institute, New York).

Banerjee, B., Kapourchali, M. H., Najnin, S., Mendel, L. L., Lee, S., Patro, C., and Pousson, M. (**2016**). "Inferring hearing loss from learned speech kernels," in *Proceedings of IEEE International Conference on Machine Learning and Applications*, pp. 26–31.

Boersma, P., and Weenink, D. (**2013**). "Praat: Doing phonetics by computer (version 5.3.51) [computer program]," http://www.praat.org (Last viewed September 1, 2013).

Cassidy, F. G. (Ed.) (**1985**). *Dictionary of American Regional English* (Belknap Press of Harvard University Press, Boston, MA).

Fairbanks, G. (**1960**). *Voice and Articulation Drillbook*, 2nd ed. (Harper and Row, New York), pp. 124–139.

Hedrick, M., Bahng, J., von Hapsburg, D., and Younger, M. S. (**2011**). "Weighting of cues for fricative place of articulation perception by children wearing cochlear implants," Inter. J. Audiol. **50**, 540–547.

Honorof, D., McCullough, J., and Somerville, B. (**2000**). "Comma gets a cure: A diagnostic passage for accent study," http://www.dialectsarchive.com/comma-gets-a-cure (Last viewed February 20, 2007).

Masterson, J. J., and Bernhardt, B. H. (**2001**). *Computerized Articulation and Phonology Evaluation System* (Pearson Education, Inc., Upper Saddle River, NJ).

Najnin, S., Banerjee, B., Mendel, L. L., Kapourchali, M. H., Dutta, J. K., Lee, S., Patro, C., and Pousson, M. (**2016**). "Identifying hearing loss from learned speech kernels," in *Proceedings of INTERSPEECH*, pp. 243–247.

Osberger, M. J., and McGarr, N. S. (**1982**). "Speech production characteristics of the hearing impaired," in *Speech and Language: Advances in Basic Research and Practice*, edited by N. Lass (Academic Press, New York), Vol. 8, pp. 227–288.

Van Riper, C. (**1963**). *Speech Correction: Principles and Methods*, 4th ed. (Prentice-Hall, Englewood Cliffs, NJ).