INTERSPEECH 2020
October 25–29, 2020, Shanghai, China

Spectral Moment and Duration of Burst of Plosives in Speech of Children with Hearing Impairment and Typically Developing Children - A Comparative Study

*Ajish K. Abraham[#1], M. Pushpavathi[#2], N. Sreedevi[#3], A. Navya[#4], C.M. Vikram[*5], and S.R. Mahadeva Prasanna[^6]*

[#]All India Institute of Speech and Hearing, Mysuru, India
[*]Indian Institute of Technology Guwahati, Guwahati, India
[^]Indian Institute of Technology Dharwad, Dharwad, India
[1]ajish68@aiishmysore.in, [2]shivanna.pushpa@gmail.com, [3]sreedeviaiish@gmail.com, [4]navyaaslp@gmail.com, [5]cmvikram@iitg.ac.in, [6]prasanna@iitg.ac.in

Abstract

Speech development in children with hearing impairment (CHI) is hampered by inadequate auditory input. Speech of CHI has reduced intelligibility compared to typically developing children (TDC), mainly because of articulatory errors. Speech language pathologists (SLPs) assess these errors through perceptual evaluation and accordingly device the protocol to correct them through several sessions of speech therapy. Automatic methods need to be developed to reduce the time and enhance the accuracy of assessment. Acoustic measures of plosives may be utilized as valuable cues for automatic assessment.

The current study was aimed to investigate the burst duration and spectral moment (centroid, skewness and kurtosis) of plosives in CHI in comparison with TDC. 24 children in the age range of 5 to 8 years, divided into group I (13 TDC) and group II (11 CHI) participated. Six words in Hindi embedded with plosives (/p/, /b/, /t̪/, /d̪/, /k/, /g/) in the initial position were used as speech material.

Burst duration, spectral centroid and skewness were found to be significantly different across the groups for most of the plosives, whereas kurtosis was not. Results indicate that these measures except kurtosis are potential cues for automatic assessment of articulatory errors.

Index Terms: burst duration, spectral moment, hearing impairment, articulatory errors

1. Introduction

Children with hearing impairment (CHI) encounter difficulties in the development of speech and language, compromising speech intelligibility to a large extent. Errors in speech production of CHI can be attributed to the degradation in the signal they hear [1]. The extent of impact can vary according to the type and degree of hearing loss. The frequent articulatory errors exhibited by CHI are substitutions, omissions, distortions and additions of phoneme.

Articulatory errors are assessed by Speech language pathologists (SLPs) through conventional method of perceptual evaluation. Based on the type and severity of errors, a protocol is formulated to correct those errors through several sessions of speech therapy. Perceptual evaluation involves time and leads to errors in assessment. Acoustic analysis permits documenting subtle variations in speech parameters which are difficult to be identified in perceptual evaluation. The development of automated assessment systems based on acoustic measures may help to improve the accuracy of the assessment and saves the time of SLP. Automated systems can also act as support systems for primary health care providers who are frontline workers in identifying persons with communication disorders. In addition, these tools can also be used by caregivers to evaluate the effectiveness of the speech therapy protocol. However, there is a lack of tools for evaluating articulation, except the tools proposed by [2] and [3]. Hence, there is a need to identify effective acoustic measures which can be used as valuable cues for automatic assessment.

Many researchers have analyzed the phonetic repertoire of speech of CHI [4, 5, 6, 7]. Stark [8] reported that such children in the early stages of speech development seem to produce the low front vowels, a neutral mid vowel or schwa, an aspirant /h/, a syllabic nasal consonant /m/ and glottal stops more often. The high vowels /i/, /u/ were infrequent in their repertoire. Children were producing front consonants /p/, /b/, /m/, /w/ more often than back consonants [7, 9] as the visibility of the front consonants makes learning easy. The most commonly reported articulatory errors are interchanging voiced/unvoiced phonemes, substituting one consonant for another, cluster reduction, abutting and omission of initial and final consonants [7, 10]. Li, Bunta, and Tomblin [11] reported that CHI (5 to 6 years of age) fail to demonstrate appropriate spectral structure of consonants.

Dorman, Kennedy, and Raphel [12] reported that burst duration is a significant acoustic cue for the place of articulation in syllable initial position for voiced plosives. Burst duration is the duration of release of puff of air or frication noise, immediately after the release of the articulatory constriction during the production of plosives. Bursts are evident for most of the plosives in the initial and medial position of the syllables, but not in final position. Thus burst duration can be a potential measure to evaluate the articulation errors in initial and medial positions.

Shape of burst spectrum is an invariant acoustic cue for identification of place of articulation [13]. Blumstein and Stevens [14] proposed the template of burst spectrum for different places of articulation of plosives. Velar stops possess compact burst spectrum with prominent peak in the range of

http://dx.doi.org/10.21437/Interspeech.2020-1805

1800-2000 Hz, but up to 4700 Hz. Alveolar stops are characterized by high-frequency spectral prominence (>4000 Hz) with diffuse raising spectrum, whereas bilabial stops are characterized by diffuse fall spectrum with spectral prominence in the low-frequency region (500-1500 Hz) [15].

To capture the spectral shape attributes of burst in plosives, spectral moments were proposed by Forrest et al. [16]. Spectral moment analysis treats the spectrum as a distribution of numbers, and measures the centroid, tilt, and peakedness of the set of numbers by using normal distribution. The spectral centroid (mean) is the average of all the numbers in the distribution [17]. Skewness of the distribution means the degree of its tilt towards the left or right with respect to the center, and kurtosis refers to the degree to which the distribution is flat or peaked [17].

98% correct classification of velar, alveolar, and bilabial stops can be achieved by evaluating the spectral moments [16]. Spectral centroid (mean), skewness, and kurtosis measures were used to characterize the voiced and unvoiced stops [18]. The voiced stops exhibit lower spectral mean due to the presence of low-frequency dominant voicing characteristics compared to unvoiced stops. Shadle and Mair [19] used spectral moments to quantify spectral characteristics of fricatives.

Earlier studies established the role of spectral moments, (centroid, skewness, and kurtosis) computed from burst spectrum, in identifying the place of articulation and voicing characteristics of plosives and fricatives. Past research motivated us to explore the possibility of using these measures to characterize the speech of CHI. The possibility to use these measures to detect and assess the articulatory errors needs to be explored. Hence, the current study was aimed to explore the potential of using these measures to assess the articulatory errors in speech of CHI.

Objectives of the study were: 1) To compare the burst duration, centroid of spectral moment, skewness and kurtosis of plosives (/p/, /t/, /k/, /b/, /d/, /g/) spoken by CHI and Typically Developing Children (TDC), 2) To investigate the significance of these measures in differentiating the speech of CHI and TDC and 3) To explore the potential of using these measures to assess the articulatory errors.

Rest of the paper is organized as follows. The materials used and procedure followed are described in Section 2. Experimental results are shown in Section 3. Section 4 provides the discussion followed by conclusion in Section 5.

2. Method

2.1. Participants

Standard group comparison research design was employed to conduct the study. 24 children, divided into group I and II, were included, utilizing a purposive convenience sampling method. Group I consisted 13 Hindi speaking TDC (6 girls and 7 boys) in the age range of 5 to 8 years (mean age - 6.8 years). They were included after an informal screening by an SLP. Those with any communication disorder, sensory or neurological problems or any other abnormalities were excluded.

Group II consisted 11 Hindi speaking CHI (5 girls and 6 boys) in the age range of 5 to 8 years (mean age - 6.9 years). These children were evaluated and diagnosed with bilateral severe to profound degree of hearing loss (sensorineural or mixed). They were fitted with digital hearing aids in both the ears. Those with any syndromic conditions or any related abnormalities were not included in the study. Written consent was given by the caregivers of the children in both the groups, prior to the study.

2.2. Materials

2.2.1. Speech Material

Bilabials (/p/, /b/), alveolar (/t/, /d/), and velars (/k/,/g/) were the speech sounds considered for the study. Six naturally-spoken meaningful words (/pəʃu:/, /bəs/, /t̪əma:t̪ər/, /ɖa:ku:/, /kəpu:r/, /ga:j/), each embedded in the initial position with one of the selected plosives, formed the speech material for the study.

2.2.2. Instrumentation

Recording of speech material was done through the microphone attached with Sound Level Meter (SLM) Type B & K 2250, using sound recording software BZ 7226. Wavesurfer (1.8.5) and MATLAB based code were used to locate and measure the parameters.

2.3. Procedure

Speech stimuli (6 Hindi words) to be presented to children, spoken by an adult female native Hindi talker, were sampled at 22 kHz (16 bit) and recorded onto a computer. Each recorded Hindi word was presented at a comfortable sound level through stereo loudspeakers to CHI and TDC after seating them inside a sound treated room, in a relaxed manner. Both loudspeakers were kept at a distance of one meter from the participant. The participants were instructed to repeat each stimulus word three times.

The Hindi words repeated three times by each child were captured by the recording set up and stored in a PC. The child was prompted to maintain a minimum interval of 500 msec between the repetitions. The recording closest to the target word was selected based on perceptual analysis by an experienced SLP, for further analysis. Speech waveforms were edited using the *Wavesurfer* (1.8.5) software. Burst onset and duration were marked for each target consonant within the word. The edited samples were run through the code written in Matlab 15a to compute the burst duration and spectral moments. To compute the spectral moments, from the burst onset, 20 msec of speech signal was selected and multiplied by Hamming window of 20 msec duration. The Fourier spectrum of Hamming windowed speech frame was computed using 512 point Fast Fourier Transform (FFT). Further, spectral centroid, skewness, and kurtosis measures were extracted from Fourier spectrum by computing the power spectrum, using the method proposed by Forrest et al. [16].

2.4. Analyses

The data collected were subjected to Shapiro Wilk normality tests and descriptive statistics. The values were compared across the groups through Mann-Whitney U test. The values within the group and across the groups were compared by means of Friedman test and post hoc analysis was done using Wilcoxon signed rank test.

3. Results

Burst duration of plosives (/p/, /t/, /k/, /b/, /d/, /g/) were measured from the recorded Hindi words of both groups.

Descriptive statistical analysis was done to compare the burst duration within the groups and across the groups. Results indicated longer burst duration for velars, followed by alveolar and bilabials in TDC, as shown in Figure 1. Burst duration in TDC ranged less than 3 msec to more than 20 msec.CHI exhibited longer burst duration for alveolar than velars and bilabials. Both groups exhibited longer burst duration for unvoiced plosives compared to voiced plosives.

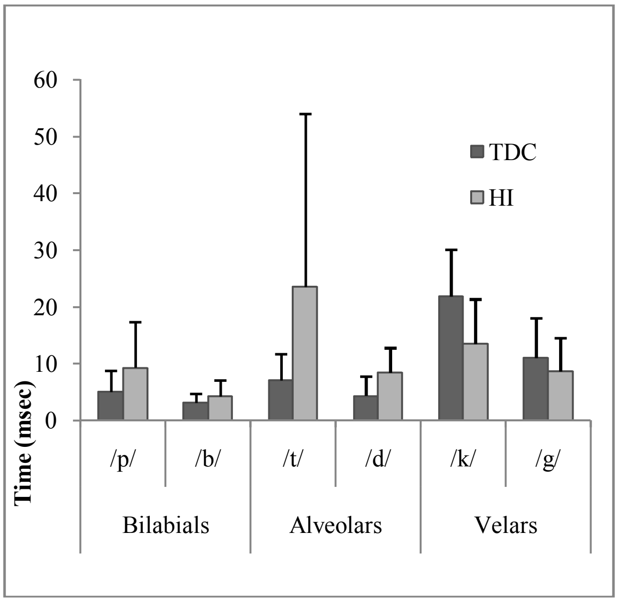

Figure 1: *Burst duration of plosives (TDC – Typically Developing Children, HI – Children with Hearing Impairment).*

Centroid of spectral moment (mean), skewness and kurtosis of plosives (/p/, /t/, /k/, /b/, /d/, /g/) were estimated. Results of descriptive statistical analysis are shown in Figure 2, 3 and 4.

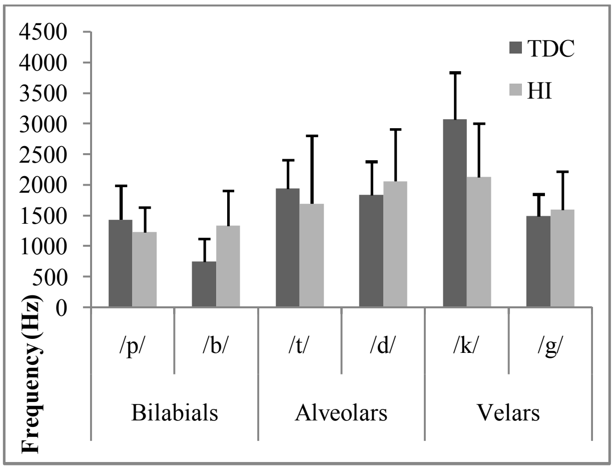

Figure 2: *Spectral Centroid of plosives (TDC – Typically Developing Children, HI – Children with Hearing Impairment).*

Spectral centroid is observed to be higher in TDC compared to CHI for all unvoiced plosives, as shown in Figure 2. For all voiced plosives, spectral centroid was observed to be lower in TDC compared to CHI.

The skewness is higher in voiced plosives than unvoiced in TDC. Similar pattern was not observed in CHI, except for velars, as evident from Figure 3. Figure 4 shows that Kurtosis measures did not exhibit any specific pattern in both groups.

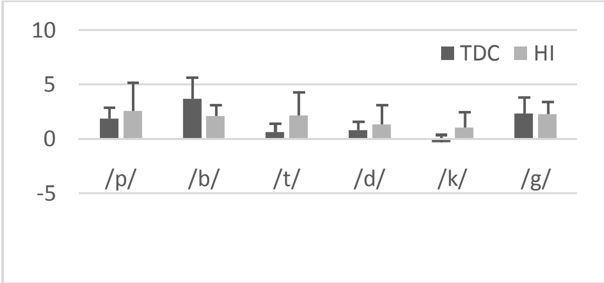

Figure 3: *Skewness of plosives (TDC – Typically Developing Children, HI – Children with Hearing Impairment).*

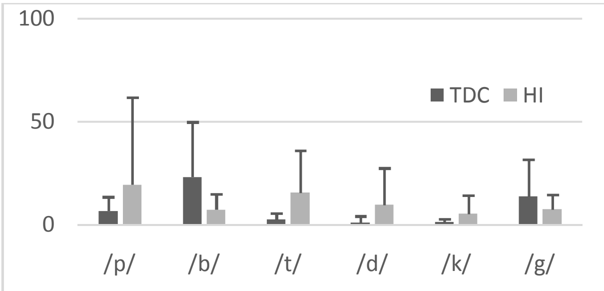

Figure 4: *Kurtosis of plosives (TDC – Typically Developing Children, HI – Children with Hearing Impairment).*

Shapiro Wilk normality test indicated the non normal distribution ($p<0.01$) of data. The data was subjected to Mann-Whitney U test to compute the effect of gender on the measured parameters in each group. Mann-Whitney U test in both groups indicated no significant effect of gender on the burst duration and spectral moments ($p>0.05$). Hence for further analysis, gender wise data was combined.

Table 1: *Results of Mann Whitney test (BD- Burst Duration, C-Centroid, S- Skewness and K-Kurtosis).*

Plosive	Variables	BD	C	S	K
/p/	MWU	49	62	78	75
	Z	-1.57	-0.87	0.00	-0.16
	P	0.11	0.38	1.00	0.87
/t/	MWU	33	38	34	31
	Z	-1.98	-1.67	-1.92	-2.10
	P	0.04	0.09	0.05	0.03
/k/	MWU	32	30	29	69
	Z	-2.28	-2.40	-2.46	-0.14
	P	0.02	0.01	0.01	0.88
/b/	MWU	60	25	35	34
	Z	-0.66	-2.69	-2.11	-2.17
	P	0.50	0.00	0.03	0.03
/d/	MWU	25	52	53	65
	Z	-2.48	-0.80	-0.74	0.00
	P	0.01	0.42	0.45	1.00
/g/	MWU	58	70	66	54
	Z	-0.78	-0.08	-0.31	-1.01
	P	0.43	0.93	0.75	0.31

Results of Mann-Whitney U test across the groups indicated significant effect of group differences on some of the measures, as shown in Table 1. Post-hoc analysis using Wilcoxon signed rank test revealed that, among the four measures, a minimum of three measures exhibited significant differences across the groups for unvoiced plosives /t/, /k/, and voiced plosive /b/. Only the burst duration of /d/ was significantly different between the groups and the plosives /p/ and /g/ did not exhibit significant differences in any of the measures between the groups.

Results indicated that the burst duration of plosives /t/, /k/, and /d/ as well as spectral centroid and skewness of plosives /t/, /k/, and /b/ could significantly discriminate the groups. However, the kurtosis of only two plosives /t/ and /b/ were significantly different between the groups.

4. Discussion

The current study investigated the burst duration and spectral moments of plosives in the speech of CHI and TDC. Earlier studies have shown burst duration as one of the robust temporal features in differentiating the plosives. Results of the current study indicated that the burst duration increased as the place of articulation moved towards posterior from anterior. The velars exhibited longer burst duration than alveolar and bilabials in TDC. Longer burst duration in velars and shorter burst duration in bilabials can be attributed to the variation in the pace of release of articulators. During the production of plosives, a brief transient burst of frication noise is generated. Bilabials are produced with the rapid movement of the lips, whereas the velar production required relatively slower movement of the tongue dorsum and mandible. Hence, the more abrupt release occurs in bilabials, slowest for velars, and in between for alveolar [20]. Similar pattern was observed in TDC, in the present study.

CHI exhibited shorter burst duration for bilabials and velars compared to alveolar especially during the production of unvoiced phonemes /t/ and /k/. Variation of burst duration from 4 msec to 30 msec was noticed in velars. CHI often produce anterior phonemes correctly than posterior phonemes. This could be due to the better visibility of the articulatory gestures augmenting the auditory cues of the phonemes [7]. Nober [21] reported that the correctly articulated plosives by CHI are observed to be 59% of bilabials, 48% of dentals, 23% of alveolar, 18% of palatals 12% of velars. Similar pattern of correct production have been reported by Smith [22] and Gold [23]. Longer burst duration of alveolar compared to the velars in CHI indicates the incorrect production of plosives.

The study investigated the significance of the measures of spectral moment in differentiating the speech of CHI and TDC. In TDC, the spectral centroid of unvoiced plosives was observed to be higher than their voiced counterparts. The spectral centroids at low frequencies in the voiced plosives could be due to the dominant low frequency energy of the voicing feature [18]. The spectral centroid measures were in the same range for bilabials, lower in alveolar and higher in velars, when compared with the reported range in the earlier studies [13, 14]. The skewness and kurtosis were high in bilabials compared to alveolar and velars in both groups except for /g/. Also these measures were high in voiced plosives, compared to unvoiced. However, negative skewness is observed for the production of /k/ in TDC, whereas all other plosives were exhibiting positive skewness. Similar results were reported by Lousada et al. [24]. This distinction in the skewness and kurtosis measures were relatively less evident in CHI, compared to TDC. The possible reason could be due to the overlap in the place of articulation of various plosives exhibited by CHI, as they tend to substitute front consonants in place of back consonants. Some of the researchers also reported that CHI fail to coordinate respiration, phonation, and articulation in their effort to produce the voicing contrasts [25]. The variations in the burst spectral measures of the current study when compared with the earlier studies [13, 14] could be attributed to the methodological differences. Current study was on children speaking Hindi whereas the referred studies were done in adults speaking English.

The study also explored the potential of using the measures of burst duration and spectral moments of plosives to assess the articulatory errors. The shorter burst duration in velars in CHI indicates shift in place of articulation to alveolar or bilabials from velar. Some children did not produce the burst at all, indicating omission of the stop consonant. This could be due to the shift in place of articulation of alveolar and velars [7]. For instance, in the current study, the burst duration of /k/ produced by the TDC is above 20 msec, whereas many CHI exhibited burst duration of less than 12 msec for /k/. This indicates substitution of a consonant produced with the anterior place of articulation similar to /t/. Another observation was that, a few CHI did not exhibit burst for the plosives in some words. The omission of the stop consonant from the word was evident on perception.

Variations in the burst spectral measures are potential cues in identifying the shift in the place of articulation. The variations in the spectral moment centroid from 3 kHz in TDC to 2 kHz in CHI observed in /k/ can cause possible distortion in perception. For example the word /ɡa:j/ was produced as /a:j/ by some CHI. Each of the articulatory errors, substitution, omission or distortion was indicated by variations in the acoustic measures.

5. Conclusions

Among the measures investigated, burst duration, spectral centroid and skewness were found to be significantly different between CHI and TDC, for some of the plosives. Kurtosis was not significantly different for most of the stimuli. Thus the study indicated that the measures - burst duration and spectral moment (centroid and skewness) of plosives - can be effectively used for automated assessment of articulatory errors in CHI. Common articulatory errors, reported in the speech of CHI - substitution, omission or distortion - were indicated by a variation in the acoustic measures investigated in the present study. These measures need to be evaluated on a larger sample to draw conclusions on identification of articulatory errors based on these measures. Also, the measures can be used for automated evaluation of articulation errors.

6. Acknowledgements

The authors acknowledge the financial support under Impacting Research Innovation and Technology (IMPRINT) scheme provided by Indian Council of Medical Research and the Ministry of Human Resource Development, Government of India.

7. References

[1] J. H. Lowenstein and S. Nittrouer, "Perception-Production Links in Children's Speech," *Journal of Speech, Language, and Hearing Research*, vol. 62, no. 4, pp. 853–867, 2019.

[2] B. H. Story and K. Bunton, "Vowel space density as an indicator of speech performance," *The Journal of the Acoustical Society of America,* vol. 141, no. 5, pp. EL458–EL464, 2017.

[3] R.van Son, C. Middag, and K. Demuynck, "Vowel Space as a Tool to Evaluate Articulation Problems," *Proc. Interspeech 2018*, pp. 357-361, 2018.

[4] A. C. D. O. Sobreira, B. M. Capo, T. S. D. Santos, and D. Gil, "Speech and language development in hearing impairment: Two-case report," *Revista CEFAC*, vol.17, no.1, pp. 308-317, 2015.

[5] M. P. Mavilya, "Spontaneous vocalization and babbling in hearing impaired infants," In G. Fant (Ed.), *International symposium on speech communication abilities and profound deafness. Washington, D.C.: A. G. Bell Association*, 1968.

[6] J. West and J. Weber, "A phonological analysis of the spontaneous language of a four year old, hard of hearing child," *Journal of Speech and Hearing Disorders*, vol. 38, no. 1, pp. 25-35, 1973.

[7] M. J. Osberger and N. S. McGarr, "Speech production characteristics of the hearing impaired," *Speech and Language,* vol. 8, pp. 221-283, 1982.

[8] R. E. Stark, "Vocalizations of the pre-school deaf child," *Progress Report. Baltimore, Md.: The Johns Hopkins School of Medicine,* 1967.

[9] R. Lach, D. Ling, A. Ling, and N. Ship, "Early speech development in deaf infants," *American Annals of the Deaf,* vol.115, no. 5, pp. 522-526, 1970.

[10] H. Levitt, C. R. Smith, and H. Stromberg, "Acoustic, articulatory, and perceptual characteristics of the speech of deaf children," In G. Fant (Ed.), *Proceedings of the Speech Communication Seminar*, Stockholm, pp. 129-139, 1974.

[11] F. Li, F. Bunta, and J.B. Tomblin, "Alveolar and post alveolar voiceless fricative and affricate productions of Spanish–English bilingual children with cochlear implants," *Journal of Speech, Language, and Hearing Research*, vol. 60, no. 9,pp. 2427–2441, 2017.

[12] M. F. Dorman, M. Studdert-Kennedy, and L. J. Raphael, "Stop consonant recognition: Release bursts and formant transitions as functionally equivalent context-dependent cues," *Perception & Psychophysics*, vol. 22, no. 2, pp. 109–122, 1977.

[13] K. N. Stevens and S. E. Blumstein, "Invariant cues for place of articulation in stop consonants," *The Journal of the Acoustical Society of America*, vol. 64, no. 5, pp. 1358-1368, 1978.

[14] S. E. Blumstein and K. N. Stevens, "Acoustic invariance in speech production: evidence from measurements of the spectral characteristics of stop consonants," *The Journal of the Acoustical Society of America*, vol. 664, no. 4, pp. 1001-17, 1979.

[15] J. Coleman, Acoustic structure of consonants [Online]. Available: http://www.phon.ox.ac.uk/jcoleman/consonant_acoustics.htm.

[16] K. Forrest, G. Weismer, P. Milenkovic, and R. N. Dougall, "Statistical analysis of word initial unvoiced obstruents: preliminary data," *The Journal of the Acoustical Society of America*, vol. 84, no. 1, pp. 115-123, 1988.

[17] T. S. Kay, *Spectral analysis of stop consonants in individuals with dysarthria secondary to stroke.* LSU Master's Theses, 2632, 2012.

[18] E. Chodroff and C. Wilson, "Burst spectrum as a cue for the stop voicing contrast in American English," *The Journal of the Acoustical Society of America*, vol. 136, no. 5, pp. 2762-2772, 2014.

[19] C. H. Shadle and S. J. Mair, "Quantifying spectral characteristics of fricatives," *Proceedings of the International Conference on Spoken Language Processing*; Philadelphia, pp. 1521–1524, 1996.

[20] K. N. Stevens, S. Y. Manuel, and M. Matthies, "Revisiting place of articulation measures for stop consonants: Implications for models of consonant production," *Proceedings of the International Congress of Phonetic Sciences*, pp. 1117-1120, 1999.

[21] H. Nober, "Articulation of the deaf," *Exceptional Children*, vol. 33, no. 9, pp. 611-621, 1967.

[22] C. R. Smith, "Residual hearing and speech production in deaf children," *Journal of Speech and Hearing Research*, vol. 18, no. 4, pp. 795-811, 1975.

[23] T. Gold, *Speech and hearing skills: comparison between hard-of-hearing and deaf children,* Unpublished doctoral dissertation, City University of New York, 1978.

[24] M. Lousada, L. M. Jesus, S. Capelas, C. Margaça, D. Simões, A. Valente, A. Hall, and V. L. Joffe, "Phonological and articulation treatment approaches in Portuguese children with speech and language impairments: A randomized controlled intervention study," *International Journal of Language & Communication Disorders,* vol. 48, no. 2, pp. 172-87, 2013.

[25] R. Whitehead and S. Barefoot, "Some aerodynamic characteristics of plosive consonants produced by hearing-impaired speakers," *American Annals of the Deaf*, vol. 125, no. 3, pp. 366-373, 1980.