Analysis and Detection of Stop Consonant Production Errors in Cleft Lip and Palate Speech

C. M. Vikram

Indian Institute of Technology Guwahati, Guwahati, India

cmvikram@iitg.ac.in

Abstract

The presence of velopharyngeal dysfunction and abnormal oral structure in individuals with cleft lip and palate (CLP) will deviate the production mechanism of stop consonants from the normal. The present thesis aims to develop an acoustic analysis based system for the detection of stop consonant production errors in children with CLP. The consonant-vowel (CV) transitions are considered as the crucial cues for the analysis of articulation errors. An event-based hierarchical algorithm is proposed for the detection of CV transitions. To analyze the epoch-to-epoch variation of spectral characteristics in the CV transitions, an epoch extraction algorithm is developed. Alternative to short-time window based time-frequency representation, single pole filter (SPF) based approach is used to epoch synchronous analysis of CV transitions. Further, the thesis aims to develop an automatic stop consonant error detection system using segmented CV transitions and SPF based features.

Indexterms: Articulation errors, cleft lip and palate, spectro-temporal analysis, transition regions.

1. Introduction

Cleft lip and palate (CLP) is a craniofacial anomaly and common most birth defect [1]. Cleft palate with or without cleft lip occurs 1 in 781 live births [2]. Individuals with cleft lip and palate (CLP) exhibit deviant characteristics of stop consonants due to the presence of inadequate functioning of velopharyngeal port, anatomical abnormalities, and mislearned articulation [1, 3]. Weak, nasalized, glottal, palatalized, pharyngeal stop substitutions, and devoicing errors are the important stop consonant errors listed in [4]. The type of speech therapy and surgery required to correct the stop consonant production errors in CLP, require a detailed evaluation of articulation errors. Acoustic analysis of articulation disorders is considered as a reliable, objective, cost-effective technique [5], when compared to perceptual and instrumental based approaches.

1.1. Related work

Compared to the detection of hypernasality in CLP [6, 7], the evaluation of articulation errors is very rarely addressed in the literature. In [5], automatic detection of glottal stops, pharyngeal backing, weak stops, and nasalized consonants using hidden Markov model (HMM) based segmentation and MFCCs has been proposed. The for The automatic detection of glottal stops using MFCCs, gammatone filter bank energies, formants estimated using linear prediction (LP) analysis, wavelet coefficient feature, and K-mean cluster classifier is proposed recently in [8].

1.2. Issues related to analysis of stop consonant errors

In the process of the assessment of articulation errors, response of the speaker for the word/sentence level stimuli containing the target phoneme is recorded [4]. Accurate segmentation of speaker's response for the target phoneme is a crucial stage in the automatic detection of articulation errors. In [5], HMM-based force alignment method is used for the segmentation. Evidence derived from short-time energy and periodicity profile is used for the segmentation of initial and final consonants in [8]. Stop consonants are very weak energy signals, whose spectro-temporal characteristics vary dynamically. Therefore, HMM and energy based methods may not be suitable for the segmentation of stop consonant errors. Segmentation and analysis of stop consonants involve the identification of closure-burst transitions [9]. In case of weak or nasalized stops, due to the inadequate development of intra-oral pressure, the burst event may not be evident [10]. In case of glottal stops, devoicing errors and nasal substitution errors, the manner of stop consonants is completely deviated due to the abnormal behavior of glottis. Hence, unlike normal stops, the articulation errors do not possess the closure-burst transitions. Alternative to burst, formant dynamics of consonant-vowel (CV) transition considered to carry the information about the place and manner of articulation of consonants [11]. Hence, in this work, CV transitions are used as appropriate regions for the analysis of articulation errors.

Conventional block processing based features such as MFCCs, LP based formants, which do not capture spectro-temporal dynamics associated with the CV transitions [12]. In CLP speech disorders, the evaluation mainly is focused on the children speech [1]. Compared to adult speech, the formant variations are highly dynamic in nature for children case, due to the rapid movement of articulators. Therefore, the analysis of such short durational CV transitions still poses a challenge for the evaluation of stop consonant production errors in CLP speech. Alternative to frame-based approaches, estimation of vocal tract spectrum by positioning the short-time analysis window around the glottal closure instant or epoch is proposed in [12]. However, estimation of epoch-synchronous vocal tract spectrum requires accurate detection of epochs. Existing epoch extraction algorithms uses LP analysis for the precise detection of epochs. Due to the presence of high pitch in children speech, LP-based methods may not perform well. Non-LP methods such as zero-frequency filter, single pole filter based approaches [13, 14] will not estimate the epochs accurately. Therefore, an accurate epoch extraction algorithm is required for the epoch synchronous estimation of the vocal tract spectrum for the analysis of CV transitions in children speech.

2. Thesis objectives

By addressing the issues related to the analysis of stop consonant errors in children with CLP, in the present thesis, the following objectives are formulated.

- Development of a knowledge-based hierarchical segmentation algorithm for the segmentation of stop consonant production errors in CLP.
- Analysis of glottal activity features for the detection of the deviant glottal source mechanism associated with the

stop consonant production errors in CLP.

- To propose an epoch extraction algorithm for the accurate detection of epochs from pathological children speech.
- To analyze the CV transitions using epoch synchronous analysis of single pole filter (SPF) based time-frequency representation (TFR).
- Demonstration of the systems for the classification of normal stop vs. misarticulated stop, normal voiced stop vs. nasalized voiced stop, and normal unvoiced stop vs. glottal stops using knowledge-based segmentation and features extraction from SPF based TFR.

3. Results from the completed work

The database consists of 31 children with repaired CLP of age range 6-12 years and 26 controlled normal (CN) children of the same age range as CLP. Velar (/k/, /g/), dental (/t/, /d/), retroflex (/T/, /D/), and bilabial (/p/, /b/) stop consonants are considered for the study. Speech material consists of consonant-vowel-consonant-vowel (CVCV) words formed by the combination of stops and vowel /a/ (example: [kaka], [gaga]). In total 1200 words from CLP and 1000 words from CN group are recorded. The recorded samples are evaluated by three expert SLPs using consonant error evaluation protocol mentioned in [4].

3.1. Hierarchical CV transition segmentation algorithm

The word stimuli consist of CV syllables containing target stop consonants. The response of the CLP speaker for the target syllable may include the distortion or substitution error, where the information about the stop consonant may be completely absent. However, we found that, within a syllable, the information about the vowel is still preserved. Using the evidence derived from glottal activity and band-pass filtered signals (500-4000 Hz), first, the locations of the syllable nuclei in CVCV words are detected. Further, using periodicity measures and band-pass filtered signals, the vowel onset points are detected. The analysis of burst evidence [9] around the vowel onset is carried out to identify the burst. If the burst is present, then the CV transition region anchored around burst onset is analyzed; otherwise, vowel onset is considered for the extraction of CV transition regions.

3.2. Detection of glottal activity errors

The presence/absence of glottal vibrations during the production of unvoiced/voiced stops is referred as the glottal activity error (GAE). Due to the inadequate development of intra-oral pressure, speakers with CLP alters the glottal activity mechanism during the production of stop consonants. The acoustic-phonetic and production based knowledge of stop consonants are exploited and an algorithm for the automatic detection of GAE is proposed in [15]. The algorithm uses zero frequency filtered and band-pass (500-4000 Hz) filtered speech signals to identify the syllable nuclei positions, followed by the detection of glottal activity characteristics of consonant present within the syllable. Based on the identified glottal activity characteristics of the consonant and $a\ priori$ voicing information of target stop consonant, the presence or absence of GAE is detected. The detected GAEs are evaluated against the ground truth derived from PRAAT-based waveform analysis.

3.3. SPF based time-frequency representation

Single pole filter (SPF) based time-frequency representation is used for the analysis of CV transitions. SPF is a first order infinite-impulse-response filter, whose impulse response is exponentially decaying in nature. Computation of STFT using SPF is termed as exponentially forgetting transform (EFT) [16]. The sharp leading edge of EFT analysis window allows the accurate detection of signal beginning. SPF has zero phase lag at the center frequency and causal impulse response. Therefore, no delay is introduced in the signal analysis. Time-frequency uncertainty i.e time-bandwidth product for EFT is $\frac{\pi}{2}$, whereas π for the Gaussian window of same length [16]. Application of SPF for the extraction of impulse-like events, i.e., glottal closure instants or epochs is proposed in [14]. Fig. 1(a)-(c) and (d)-(f) represent the speech waveform, STFT, and SPF-based spectrograms, respectively, for the CV transitions of normal velar stop (/k/) and glottal stop, respectively. Formant transitions are better visualized in SPF-based spectrogram than STFT.

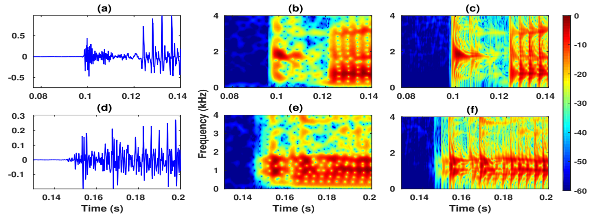

Figure 1: *Illustration of STFT and SPF based spectrograms. (a)-(c) and (d)-(f), represent speech waveform, STFT and SPF based spectrograms for velar stop and glottal stop, respectively.*

3.4. Epoch extraction from pathological children speech

The vertical striations present in the spectrogram due to the abrupt glottal closure are highly localized to time in the SPF-based spectrograms (Fig. 1(c) and(f)). Further, multi-scale product of each SPF-based filtered envelope is computed to enhance the transition present around the epoch. The impulse-like sequence derived using the SPF based envelopes and the multi-scale product is used as the pre-processed signal for the epoch extraction. A short reference interval defined around the positive zero crossings of ZFFS is used to locate the epoch from the pre-processed signal. The SPF and ZFFS based epoch extraction algorithm showed better performance for the pathological children speech database [17], under clean as well as noisy conditions, when compared to LP-based approaches [18].

4. Conclusion and future directions

Analysis of the various stop consonant production errors revealed that the vowel landmarks are not distorted in CLP. Hence vowel onset points can be used as anchor points for the segmentation of CV transitions. Alternative to STFT, SPF-based TFR found to be better for the analysis of CV transitions. To analyze the epoch-to-epoch variation of vocal tract features of children speech, an epoch extraction algorithm is proposed. Future work includes the extraction of features from SPF-based TFR for the analysis of stop consonant production errors in CLP. Further, using segmented CV transitions and SPF based features, a system for the automatic detection of stop consonant production errors need to be developed.

5. Acknowledgement

I express sincere gratitude to my supervisor Prof. S R Mahadeva Prasanna for guiding me throughout my PhD. Also, I would like to thank Prof. Ajish K. Abraham and Prof. M. Pushpavathi of All India Institute of Speech and Hearing, Mysore for their help and support for the data collection. This work is in part supported by the project grants, for the projects entitled "NASOSPEECH" funded by the Department of Biotechnology (DBT), Govt. of India and "ARTICULATE +" funded by the Ministry of Human Resource Development (MHRD), Govt. of India.

6. References

[1] A. W. Kummer, *Cleft Palate & Craniofacial Anomalies: Effects on Speech and Resonance*. Cengage Learning, 2013.

[2] S. Raman, M. Jacob, M. S. Jacob, and R. Nagarajan, "Providing intervention services for communication deficits associated with cleft lip and/or palatea retrospective analysis," *Asia Pac Disabil Rehabil J*, vol. 15, pp. 122–9, 2004.

[3] D. Sell, A. Harding, and P. Grunwell, "A screening assessment of cleft palate speech (great ormond street speech assessment)," *International Journal of Language & Communication Disorders*, vol. 29, no. 1, pp. 1–15, 1994.

[4] G. Henningsson, D. P. Kuehn, D. Sell, T. Sweeney, J. E. Trost-Cardamone, and T. L. Whitehill, "Universal parameters for reporting speech outcomes in individuals with cleft palate," *The Cleft Palate-Craniofacial Journal*, vol. 45, no. 1, pp. 1–17, 2008.

[5] A. Maier, F. Hönig, T. Bocklet, E. Nöth, F. Stelzle, E. Nkenke, and M. Schuster, "Automatic detection of articulation disorders in children with cleft lip and palate," *The Journal of the Acoustical Society of America*, vol. 126, no. 5, pp. 2589–2602, 2009.

[6] P. Vijayalakshmi, M. R. Reddy, and D. O'Shaughnessy, "Acoustic analysis and detection of hypernasality using a group delay function," *IEEE Transactions on biomedical engineering*, vol. 54, no. 4, pp. 621–629, 2007.

[7] M. Golabbakhsh, F. Abnavi, M. Kadkhodaei Elyaderani, F. Derakhshandeh, F. Khanlar, P. Rong, and D. P. Kuehn, "Automatic identification of hypernasality in normal and cleft lip and palate patients with acoustic analysis of speech," *The Journal of the Acoustical Society of America*, vol. 141, no. 2, pp. 929–935, 2017.

[8] L. He, J. Zhang, Q. Liu, J. Zhang, H. Yin, and M. Lech, "Automatic detection of glottal stop in cleft palate speech," *Biomedical Signal Processing and Control*, vol. 39, pp. 230–236, 2018.

[9] T. Ananthapadmanabha, A. Prathosh, and A. Ramakrishnan, "Detection of the closure-burst transitions of stops and affricates in continuous speech using the plosion index," *The Journal of the Acoustical Society of America*, vol. 135, no. 1, pp. 460–471, 2014.

[10] B. J. Philips and R. D. Kent, "Acousticphonetic descriptions of speech production in speakers with cleft palate and other velopharyngeal disorders," *Speech and Language*, vol. 11, pp. 113 – 168, 1984.

[11] D. J. Sharf and T. Hemeyer, "Identification of place of consonant articulation from vowel formant transitions," *The Journal of the Acoustical Society of America*, vol. 51, no. 2B, pp. 652–658, 1972.

[12] B. Yegnanarayana and R. N. Veldhuis, "Extraction of vocal-tract system characteristics from speech signals," *IEEE transactions on Speech and Audio Processing*, vol. 6, no. 4, pp. 313–327, 1998.

[13] K. S. R. Murty and B. Yegnanarayana, "Epoch extraction from speech signals," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 16, no. 8, pp. 1602–1613, 2008.

[14] C. M. Vikram and S. R. Mahadeva Prasanna, "Epoch extraction from telephone quality speech using single pole filter," *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP)*, vol. 25, no. 3, pp. 624–636, 2017.

[15] C. M. Vikram, S. R. Mahadeva Prasanna, A. K. Abraham, M. Pushpavathi, and K. S. Girish, "Detection of glottal activity errors in production of stop consonants in children with repaired cleft lip and palate," *Proc. Interspeech (Accepted)*, 2018.

[16] S. Tomazic, "On short-time fourier transform with single-sided exponential window," *Signal processing*, vol. 55, no. 2, pp. 141–148, 1996.

[17] B. Woldert-Jokisz, "Saarbruecken voice database," 2007.

[18] C. M. Vikram and S. R. Mahadeva Prasanna, "Epoch extraction from pathological children speech using single pole filtering approach," *Proc. Interspeech (Accepted)*, pp. 449–453, 2018.