INTERSPEECH 2019
September 15–19, 2019, Graz, Austria

Modification of Devoicing Error in Cleft Lip and Palate Speech

Protima Nomo Sudro[1], *S R Mahadeva Prasanna*[1,2]

[1]Indian institute of Technology Guwahati, Guwahati, India
[2]Indian institute of Technology Dharwad, Dharwad, India
(protima,prasanna)@iitg.ernet.in

Abstract

The cleft of the lip and palate (CLP) caused by structural and functional deformation leads to various speech-related disorders, which substantially degrades the speech intelligibility. In this work, devoiced stop consonants in CLP speech are analyzed, and an approach is proposed for its modification in order to enhance the speech intelligibility. The devoicing errors are primarily characterized by the absence of voicebar in the closure interval and relatively longer voice onset time (VOT). The proposed approach first segments the regions corresponding to the closure interval and VOT based on the knowledge of glottal activity, voice onset point, voice offset point, and burst onset point. In the next stage, specific transformations are performed for the modification of closure bar and VOT respectively. For transformation, first different transformation matrices are learned for closure bar and VOT from normal and CLP speakers. The transformation matrix is optimized using nonnegative matrix factorization method. Further, the corresponding transformation matrices are used to modify the closure bar and VOT separately. The subjective evaluation results indicate that the devoiced stop consonants tend to exhibit the characteristics of voiced stop consonants.

Index Terms: Devoicing error, CLP speech, intelligibility enhancement, nonnegative matrix factorization.

1. Introduction

The intelligibility of CLP speech is mostly affected due to deviant articulatory features, and it often leads to communicative impairments [1]. Available data indicate that cleft occurs in around $45,193$ in $30,665,615$ live births worldwide [2]. Individuals with CLP, regardless of whether the articulator is functioning at any given point of time, have a history of velopharyngeal dysfunction (VPD). The associated VPD leads to the risk of developing speech production errors. The deviant speech characteristics associated with CLP is grouped into four categories: hypernasality, compensatory articulation, nasal air emission, and voice disorders [3, 4]. The correction of speech production errors in CLP speech requires a long period of time. Clinically, the improvement of speech intelligibility can be achieved through surgery, prosthesis, and therapy. The structural correction done by surgery might not result in functional correction due to which, deviant speech persists even after surgery. Speech therapy is mostly recommended by speech-language pathologists (SLPs) to correct the functional disorders [5].

The main aim of speech therapy is to enhance speech intelligibility. Generally, SLPs consider certain assessment techniques to identify the characteristics of a speech disorder and employ appropriate therapy technique for individuals with CLP [6]. For the proper insight about the production & perception of deviant phoneme and target phoneme, auditory discrimination test is carried out. The auditory discrimination test is considered one of the regular assessment technique in speech therapy. Usually, an SLP creates awareness of the disorder by simulating the deviant speech sound and presenting it to the individual along with correct speech sounds. Sometimes SLPs also used various other techniques like phonetic placement technique, biofeedback (audio, visual, tactile), speech sound perception training, etc. during the speech therapy [6, 7, 8]. In several researches apart from CLP speech, it has been suggested that presenting the individuals with their own deviant speech increases the awareness of disorder and tend to simulate self-monitoring [9, 10, 11, 12]. Along with SLPs simulation, providing the CLP speakers with acoustically modified deviant speech as auditory feedback may be effective in increasing awareness and facilitating speech production learning provided they have an understanding of correct production mechanism. It will also motivate the individuals with CLP by giving a preview of what the voice would sound like after successful speech therapy.

Speech sound disorders include a variety of aspects that disrupts the ability to communicate in demanding situations. To help the individuals with speech disorders, intelligibility enhancement of different types of pathological speech based on signal processing method have been reported in the literature. Dysarthric speech enhancement techniques include spectral modifications based on Gaussian mixture mapping, modification of formants F1 and F2, correction of devoiced stop consonant, improving quality of continuous speech, and acoustic transformation [13, 14, 15, 16]. Various methods are employed for enhancing alaryngeal speech namely, modified spectral subtraction, reducing spectral distortion by formant enhancement using chirp z-transform & cepstral weighting, statistical voice conversion (VC) technique, imposing artificial contour on speech signals [17, 18]. Articulation disorder resulting from athetoid cerebral palsy, oral surgery, wide glossectomy/segmental mandibulectomy also impair speech intelligibility in large extent. The intelligibility of such an articulation disordered speech is improved using non-negative matrix factorization (NMF) based VC technique, spectral conversion while preserving individuality, and GMM based VC [19, 20].

Improvement in the speech intelligibility of the various speech disorders is executed to help them communicate easily. However, intelligibility enhancement of CLP speech is not studied abundantly except for two recent works [21, 22], which addressed misarticulated fricative /s/ modification and hypernasal speech modification. These studies overlooked issues like same acoustic transformation applied over the entire utterance and manual segmentation of speech signals, thus making the studies unfeasible for phoneme-specific real-time modifications. Therefore, motivated by the importance of intelligibility enhancement of CLP speech and by the pathological speech enhancement works described above, we intend to modify the CLP speech intelligibility. Specifically, modification of devoicing error where voiced stops are perceived as their respective voiceless cognates in CLP speech is attempted in the

http://dx.doi.org/10.21437/Interspeech.2019-2604

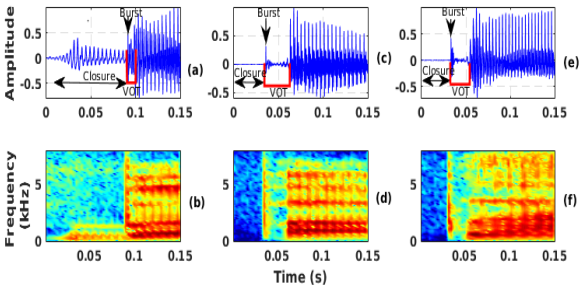

Figure 1: *Waveform and spectrogram of the (a-b) syllable /ba/ & (c-d) syllable /pa/ of normal speaker and (e-f) syllable /ba/ of CLP speaker which tends to be /pa/.*

current work. The characteristics of stops are determined by the acoustic events: onset of closure, closure interval, and burst onset. The closure interval denotes a state when the articulators are held together, completely obstructing the airflow creating a pressure build up behind the constriction and burst onset is an event when the articulators are set apart resulting in burst generation. The syllable /ba/ and /pa/ of normal speaker and syllable /ba/ of CLP speaker is shown in Fig. 1. The difference between the voiced and devoiced stop /b/ in Fig. 1(a-b) & (e-f) imply the absence of low-frequency component generated by the quasi-periodic excitation of the vocal folds. Like the normal speaker's voiceless stop /p/ shown in Fig. 1(c-d), the noise burst is preceded by silence in the closure interval of devoiced stop /b/ in Fig. 1(e-f).

As stop consonants exhibit dynamic spectro-temporal behaviour corresponding to the acoustic events [23], a common transformation method may not result in intelligible speech. Hence, in the current work, specific transformations are performed for different events of stops. Before the correction of devoiced stops, the region for modification is first specified using the knowledge of burst onset point and glottal activity regions. Once after determining the regions for modification enhancement is performed corresponding to the specific events. Voicebar is transformed using the derived transformation matrix from normal voiced stops and devoiced stops of CLP speakers. If the VOT of devoiced stop exceeds the threshold, then it is subject to modification. Besides the VOT, the deviated spectral components present in the VOT are also modified.

Rest of the paper is organized as follows. Section 2 discussed speech data and analyze the characteristics of devoiced stops. The modification of devoiced stops, which include segmentation followed by transformation, is illustrated in Section 3. In Section 4, the experimental results are discussed. Finally, conclusion and future aspects are discussed in Section 5.

2. Analysis of devoiced stop consonants in CLP speech

2.1. Speech Data

Speech utterances used in this work were collected from All India Institute of Speech and Hearing (AIISH), Mysore, India. Collected data consists of 60 native Kannada speakers of which 29 (17 male and 12 female) are speakers with CLP, and 31 (12 male and 19 female) are non-CLP control speakers. The age of CLP and non-CLP participants are 9 ± 2 years (mean $\pm$ SD) and 10 ± 2 years (mean $\pm$ SD), respectively.

The collected speech database consists of vowel phonation, nonsense VCV, and CVCV words (V and C correspond to different vowels and consonants, respectively) meaningful words and sentences. In the current work, only nonsense CVCV words are used for detection and enhancement. For the 29 CLP speakers, the manifestation of speech disorders was labeled by 3 expert SLPs of AIISH. Speech samples were recorded in clean room condition using Speech Level Meter (SLM) at 48 kHz sampling rate and 16 bit resolution. The speech signals are down-sampled to 16 kHz before processing them. The database consists of speech samples that exhibit disorders like hypernasality, articulation errors, and nasal air emission (NAE). As the focus of this work is on the modification of devoicing error, we exclude speech samples which do not exhibit devoicing of stops. Table 1 shows the total number of tokens examined for the devoicing error.

Table 1: *Description of the devoiced phonemes.*

Distortion type	No. of tokens
/b/ $\rightarrow$ /p/	65
/d/ $\rightarrow$ /t/	63
/g/ $\rightarrow$ /k/	64

As voicing in stop consonant is not only the function of voicebar but also other attributes like burst amplitude and frequency, VOT duration, etc. [24]. Therefore in the current work, we consider the different events (onset of closure, closure, and burst onset) of stops for the modification. Voicing attribute in the closure interval generally termed as voicebar, is a short duration quasi-periodic signal having a dominant spectral peak around 200 Hz. The noise burst preceded by voicebar results from a sudden release of air pressure built up during the closure interval. The interval between the release of noise burst and the start of the glottal vibration is denoted as VOT and it varies with place of articulation (PoA) in stop consonants. The difference between normal voiced stops and devoiced stops in CLP speech can be observed from Table 2. It can be observed that the VOT of the devoiced stops /b/ and /d/ in CLP speech are significantly different compared to normal voiced stops. An ANOVA test implies the distinctive nature of devoiced stop relative to normal voiced stop with $p<0.001$. However, a minute difference in VOT observed between the devoiced stop and normal voiced stop /g/ is not significantly different ($p>0.001$). In case of spectral mean, devoiced stop /d/ shows maximum distinction relative to normal stop /d/ ($p<0.001$). Although, the spectral mean of devoiced /b/ and /g/ have a small difference compared to normal /b/ and /g/, but they are statistically significant ($p<0.001$).

Table 2: *Mean and standard deviation of VOT (ms) and spectral mean (Hz) of the burst of the voiced stops of normal speakers and devoiced stops of CLP speakers.*

	/b/	/d/	/g/
Normal-VOT ($\mu \pm \sigma$)	8.4±2.8	8.9±3.6	18±9.1
CLP-VOT ($\mu \pm \sigma$)	15.1±7.1	20.6±13.3	22±11.3
Normal-spectral mean ($\mu \pm \sigma$)	1528±44.3	1582±251.9	2005.7±145.8
CLP-spectral mean ($\mu \pm \sigma$)	1757±620	2511±212.9	2396.4±180.6

3. Transformation of devoiced stops

The framework for the modification of stops corresponding to the specific distorted acoustic events is illustrated in Fig 2. Prior to enhancement, the location of burst and glottal activity regions are detected from the input speech signal. The region between the detected burst location and onset point of the glottal activity is considered as voice onset time (X_{VOT}). If X_{VOT} is observed to be deviated in terms of duration and/or spectral characteristics, then it is modified using the corresponding transformation matrix $\hat{W}_{VOT}$. The transformation matrix $\hat{W}$ is obtained using the iterative solution given in [27]. For

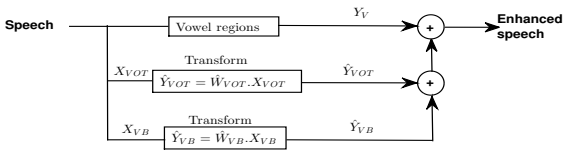

Figure 2: *Block diagram illustrating voice onset time (X_{VOT}) and voicebar (X_{VB}) transformations to obtain overall enhanced speech.*

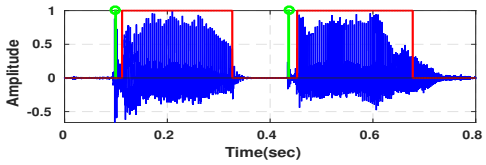

Figure 3: *Unvoiced stop consonant segmentation. Glottal activity detected (GAD) regions are marked in red and instant of burst locations are marked in green.*

word-initial condition, 60 ms before burst onset point is considered for voicebar (X_{VB}) modification and word-medial condition, the region between burst onset and previous vowel offset point is considered for X_{VB} modification. 60 ms is the average voicebar duration obtained from normal speakers voiced stops. The enhanced speech signal is then obtained by concatenating the modified voicebar ($\hat{Y}_{VB}$) and VOT ($\hat{Y}_{VOT}$) with the unprocessed segments of the speech utterance. In this work, we consider specific transformations for different events of stop consonants because each event corresponds to different spectral characteristics. The learned transformation matrix ($\hat{W}$) might be weighted by a mixture of many spectral components, which when exploited for modification may not result in enhanced speech. Therefore, we obtain specific transformation matrices for different events of the voiced stop consonant.

3.1. Segmentation algorithm

For segmentation, the burst onset point and glottal activity regions are first determined and using this information, the voicebar and VOT are specified. The segmentation of the signal begins with the detection of glottal activity region followed by calculating plosion index (PI) for capturing the abrupt increase in energy. At first, the glottal activity regions are detected using zero frequency filtering (ZFF) approach [25]. In the ZFF process, the differenced speech signal is passed through a cascade of two ideal zero Hz resonator. A process of local mean subtraction removes the cumulative DC bias present in the resonator output. The local mean subtracted signal is termed as zero frequency filtered signal (ZFFS). The positive zero crossings of the ZFFS corresponds to the glottal closure instants/epoch locations. The first order slope of ZFFS is calculated at each epoch locations and it is termed as the strength of excitation (SOE). The SOE is relatively higher for voiced regions compared to unvoiced regions. Using appropriate threshold value on SOE, the region with higher SOE values are considered as glottal activity regions.

Once the glottal activity detection (GAD) is performed, the region prior to the GAD onset point is investigated for the presence of burst. It is accomplished by exploiting PI [26] which is defined as,

$$PI(m_o, n_1, n_2) = \frac{|X(m_o)|}{X_{avg}(n_1, n_2)}$$
$$X_{avg}(n_1, n_2) = \frac{\sum_{i=m_o-(n_1+n_2)}^{i=m_o-(n_1+1)} |X(i)|}{n2} \quad (1)$$

where, m_o denotes the sample of interest, n_1 is the offset from m_o and X_{avg} denotes average of absolute amplitudes of n_2 samples. n_1 and n_2 are chosen as 6 ms and 16 ms respectively. The threshold of PI corresponds to 9 dB, which is generally used in literature for burst detection [26]. In Fig. 3 the detected bursts and glottal activity regions are shown for the devoiced /b/ in /a/ context. It can be observed that due to devoicing, the burst is preceded by a silence region which is detected using the PI metric. The detection accuracy of the burst location within 30 ms time duration, for all the three, devoiced stops: /b d g/ in vowel context /a/ are shown in Table 3. It indicates that the devoiced /g/ which sounds like /k/ is showing a minimum recognition rate of 73.91%. However, the bursts of /b/ sounding like /p/ shows a maximum recognition rate with 96% overall accuracy.

Table 3: *Detection accuracy of burst onset points for the devoiced stops of CLP speakers.*

Devoiced stops	Detection accuracy(%)
/b/ →/p/	96
/d/ →/t/	84.09
/g/ →/k/	73.91

3.2. Spectral transformation

To exploit transformations corresponding to different events of the stop consonant, first specific transformation matrices are learned, followed by multiplying it with the desired region of modification. Fig. 4, shows the process of learning a transformation matrix, $\hat{W}$ from the source and target dictionaries, D_s & D_t, which is optimized using the NMF method. To obtain the converted spectral features, the source spectral matrix is multiplied by the transformation matrix as, $\hat{Y} = \hat{W} \times X$.

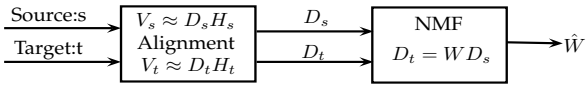

Figure 4: *Training of aligned spectral pairs to obtain transformation matrix $\hat{W}$ from normal and CLP speakers.*

The process of estimating the dictionaries first involves, representation of the spectrogram V of dimension $F \times K$ as a linear combination of basis and weights,

$$v_l = \sum_{j=1}^{J} d_j h_{jl} = DH, \text{given}, D \geq 0, h_l \geq 0 \quad (2)$$

where, v_l denote the l^{th} frame of the speech signal, $d_j h_{jl}$ denote the j^{th} basis and weight respectively, $D = [d_1, d_2, ..., d_I] \in R^{F \times I}$, where d_i is the i^{th} exemplar and $H \in R^{I \times K}$ are the dictionary and activity of the frame l respectively. F is the feature dimension, and K denotes the number of frames. D_s & D_t represent the collection of source and target basis, respectively. The source dictionary is constructed using the source features from the disordered spectrogram of CLP speech. The target dictionary is constructed using the target features attained from the spectrogram of normal speech. The two dictionaries consist of aligned magnitude spectral sequences because they are derived from the aligned speech signals obtained using dynamic time warping (DTW) method. Given the parallel spectral sequences, $A = D_s$ and $B = D_t$, the target spectral

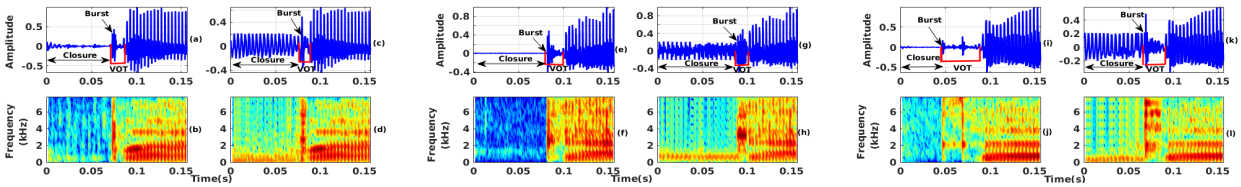

Figure 5: *Illustration of the waveform and spectrogram of (a-b) devoiced /ba/ and (c-d) modified /ba/, (e-f) devoiced /da/ and (g-h) modified /da/, (i-j) devoiced /ga/ and (k-l) modified /ga/.*

matrix B can be approximated by $\hat{W}A$ using Kullback-Leibler divergence $\mathscr{D}_{KL}$, given as

$$Z = \mathscr{D}_{KL}(B \,||\, \hat{W}A) \tag{3}$$

The approximation given in Eqn. 3 is minimized by iteratively applying the multiplicative updating rule given in [27] as follows:

$$\hat{W} \leftarrow \hat{W} \otimes \frac{\left(\frac{B}{\hat{W}A}\right)A^T}{1_{F\times K}A^T} \tag{4}$$

where, $\hat{W}$ is commonly initialized with an all-ones matrix, $\otimes$ denotes element-wise multiplication and $1 \in R^{F\times K}$, represents an all-ones matrix. An inverse short time Fourier transform is applied to the transformed spectral sequences and recombined it with the original unprocessed speech signal.

4. Experiments and results

The effectiveness of the transformation approach discussed above is examined in this section. The transformed devoiced stops are depicted in Fig 5. From Fig 5(c) and (d), the presence of voicebar is observed in the silence region of the closure interval, burst and VOT is also enhanced compared to Fig 5(a) and (b). In the enhancement process, voicebar region is first modified. Next, based on the deviant spectral characteristics of devoiced stop VOT, it is transformed. The VOT for devoiced stop of CLP speech is then aligned with that of the normal stop VOT using DTW. The warped version is replaced with the transformed VOT for devoiced stop of CLP speech, such that it temporally matches the normal voiced stop VOT. After modification, the characteristics of the transformed devoiced stop /b/ resemble with the normal speaker's voiced stop /b/ shown in Fig 1(a) and (b). Similar transformations are also observed for the devoiced stops /d/ and /g/ depicted in Fig 5(e-h) and Fig 5(i-l) respectively. The impact of modification across all the samples of three devoiced stops are shown in Fig 5. It can be observed from Fig 5 that, the characteristics of the enhanced speech is similar to normal voiced stop.

4.1. Subjective intelligibility metric

To evaluate the performance of the transformed devoiced stops in CLP speech, we conducted a perceptual evaluation test. The original and the modified speech utterances are labeled randomly to avoid any bias and presented to the listeners one at a time. All the listenings were made through headphones. Ten listeners with an understanding of speech signal processing have participated in the perceptual evaluation. The listeners were instructed to evaluate each of the utterances on a scale ranging from 0 to 100. This method of evaluation is adopted from [28]. Here, 0 represents utterance consisting of one of the voiceless stop (/p/ or /t/ or /k/) and 100 represents utterance consisting of one of the voiced stop (/b/ or /d/ or /g/). For the listener's, when they were asked to make a forced-choice in a binary setting, it becomes too coarse to be of real perceptual value. Hence, the scale here consists of a range of choices for the listener's to rate the speech files based on the attribute of voicing/devoicing.

Table 4: *Percentage of words correctly identified by each listener. O & M represents the original deviant CLP speech utterance and its modified version respectively.*

Listener's ID	/p/ → /b/		/t/ → /d/		/k/ → /g/	
	O	M	O	M	O	M
L1	15	70	0	90	0	85
L2	15	85	0	81	50	85
L3	10	75	20	76	5	55
L4	10	72	15	89	10	75
L5	5	60	50	90	0	50
L6	0	78	40	82	10	67
L7	0	68	5	79	0	80
L8	20	60	50	80	10	80
L9	0	80	50	88	10	71
L10	10	80	0	77	5	73
Average ($\mu \pm \sigma$)	8.5 ± 7.09	72.8 ± 8.45	23 ± 22.26	83.2 ± 5.5	10 ± 14.7	72.1 ± 11.9

Table 4 shows the performance of all the listeners for each of the speech utterances, averaged across the samples with same stop consonant. The transformed speech samples show values inclined towards 100, indicating that the characteristics of the devoiced stop consonants tend to be like the voiced stop consonants. A pair-wise t-test is performed between each pair of speech utterances for /p/ → /b/, /t/ → /d/ and /k/ → /g/ respectively. All pairs are observed to be significantly different with p-value < 0.001.

5. Conclusion and future work

In this work, an approach is presented for the spectral transformation of devoiced stop consonant. Prior to transformation, the specific regions for modification are segmented using the knowledge of the glottal activity and burst location. Different transformation matrices are learned for the specific acoustic events. In the transformation stage, the specific events of stops are modified using the learned transformation matrices. The illustration of modified speech signal & spectrogram and subjective evaluation results implies that the devoiced stops tend to exhibit voiced stop like characteristics. However, in real environment scenarios, the segmentation accuracy of the above mentioned acoustic events may vary, and hence the enhancement may be affected. The work presented here is an enhancement shown for phoneme-specific sound units in consonant-vowel-consonant-vowel structures. To meet the real environment scenarios, further exploration of the proposed approach is yet to be done to study the effect of reverberation and background noise, especially in meaningful words and spontaneous speech.

6. Acknowledgements

The authors would like to thank Dr. M.Pushpavathi and Dr. Ajish Abraham, AIISH Mysore, for providing insights about CLP speech disorder. The authors would like to thank the research scholars of Signal processing lab for their participation in the subjective test. This work is in part supported by a project entitled NASOSPEECH: Development of Diagnostic system for Severity Assessment of the Disordered Speech funded by the Department of Biotechnology (DBT), Govt. of India.

7. References

[1] K. Van Lierde, M. De Bodt, J. Van Borsel, F. Wuyts, and P. Van Cauwenberge, "Effect of cleft type on overall speech intelligibility and resonance," *Folia phoniatrica et logopaedica*, vol. 54, no. 3, pp. 158–168, 2002.

[2] V. Panamonta, S. Pradubwong, M. Panamonta, and B. Chowchuen, "Global birth prevalence of orofacial clefts: a systematic review," *J Med Assoc Thai*, vol. 98, no. Suppl 7, pp. S11–21, 2015.

[3] J. E. Trost, "Articulatory additions to the classical description of the speech of persons with cleft palate," *Cleft Palate J*, vol. 18, no. 3, pp. 193–203, 1981.

[4] M. Schuster, A. Maier, T. Haderlein, E. Nkenke, U. Wohlleben, F. Rosanowski, U. Eysholdt, and E. Nöth, "Evaluation of speech intelligibility for children with cleft lip and palate by means of automatic speech recognition," *International Journal of Pediatric Otorhinolaryngology*, vol. 70, no. 10, pp. 1741–1747, 2006.

[5] L. Nord and G. Ericsson, "Acoustic investigation of cleft palate speech before and after speech therapy," *Journal of STL-QPSR*, vol. 26, no. 4, pp. 15–27, 1985.

[6] A. W. Kummer, *Cleft palate & craniofacial anomalies: Effects on speech and resonance*. Nelson Education, 2013.

[7] A. Bessell, D. Sell, P. Whiting, S. Roulstone, L. Albery, M. Persson, A. Verhoeven, M. Burke, and A. R. Ness, "Speech and language therapy interventions for children with cleft palate: a systematic review," *The Cleft Palate-Craniofacial Journal*, vol. 50, no. 1, pp. 1–17, 2013.

[8] A. Jahanbin, M. R. Pahlavannezhad, M. Savadi, and N. Hasanzadeh, "The effect of speech therapy on acoustic speech characteristics of cleft lip and palate patients: a preliminary study," *Special Care in Dentistry*, vol. 34, no. 2, pp. 84–87, 2014.

[9] L. I. Shuster, "The perception of correctly and incorrectly produced/r," *Journal of Speech, Language, and Hearing Research*, vol. 41, no. 4, pp. 941–950, 1998.

[10] D. G. Jamieson and S. Rvachew, "Remediating speech production errors with sound identification training," *Journal of Speech-Language Pathology and Audiology*, vol. 16, no. 3, pp. 201–210, 1992.

[11] D. M. Shiller, S. Rvachew, and F. Brosseau-Lapré, "Importance of the auditory perceptual target to the achievement of speech production accuracy." *Canadian Journal of Speech-Language Pathology & Audiology*, vol. 34, no. 3, 2010.

[12] D. M. Shiller and M.-L. Rochon, "Auditory-perceptual learning improves speech motor adaptation in children." *Journal of Experimental Psychology: Human Perception and Performance*, vol. 40, no. 4, p. 1308, 2014.

[13] A. B. Kain, J.-P. Hosom, X. Niu, J. P. van Santen, M. Fried-Oken, and J. Staehely, "Improving the intelligibility of dysarthric speech," *Speech communication*, vol. 49, no. 9, pp. 743–759, 2007.

[14] C. Shilpa, V. Swathi, V. Karjigi, K. Pavithra, and S. Sultana, "Landmark based modification to correct distortions in dysarthric speech," in *Communication (NCC), 2016 Twenty Second National Conference on*. IEEE, 2016, pp. 1–6.

[15] A. Prakash, M. R. Reddy, and H. A. Murthy, "Improvement of continuous dysarthric speech quality," in *Proc. SLPAT 2016 Workshop on Speech and Language Processing for Assistive Technologies*, 2016, pp. 43–49.

[16] F. Rudzicz, "Adjusting dysarthric speech signals to be more intelligible," *Computer Speech & Language*, vol. 27, no. 6, pp. 1163–1177, 2013.

[17] M. Hagmüller, *Speech enhancement for disordered and substitution voices*. Citeseer, 2009.

[18] N. Bi and Y. Qi, "Application of speech conversion to alaryngeal speech enhancement," *IEEE transactions on speech and audio processing*, vol. 5, no. 2, pp. 97–105, 1997.

[19] S.-W. Fu, P.-C. Li, Y.-H. Lai, C.-C. Yang, L.-C. Hsieh, and Y. Tsao, "Joint dictionary learning-based non-negative matrix factorization for voice conversion to improve speech intelligibility after oral surgery," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 11, pp. 2584–2594, 2017.

[20] K. Tanaka, S. Hara, M. Abe, and S. Minagi, "Enhancing a glossectomy patient's speech via gmm-based voice conversion," in *Signal and Information Processing Association Annual Summit and Conference (APSIPA), 2016 Asia-Pacific*. IEEE, 2016, pp. 1–4.

[21] P. N. Sudro, S. Kalita, and S. R. M. Prasanna, "Processing transition regions of glottal stop substituted /s/ for intelligibility enhancement of cleft palate speech," in *Interspeech*, 2018.

[22] C. Vikram, N. Adiga, and S. M. Prasanna, "Spectral enhancement of cleft lip and palate speech." in *INTERSPEECH*, 2016, pp. 117–121.

[23] V. Karjigi and P. Rao, "Classification of place of articulation in unvoiced stops with spectro-temporal surface modeling," *Speech Communication*, vol. 54, no. 10, pp. 1104–1120, 2012.

[24] A. Suchato, "Classification of stop consonant place of articulation," Ph.D. dissertation, Massachusetts Institute of Technology, 2004.

[25] 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.

[26] 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.

[27] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," in *Advances in neural information processing systems*, 2001, pp. 556–562.

[28] S. Strömbergsson, "The/k/s, the/t/s, and the inbetweens: Novel approaches to examining the perceptual consequences of misarticulated speech," Ph.D. dissertation, KTH Royal Institute of Technology, 2014.