Synopsis report on

Glottal Activity Region based Processing for Speech Synthesis

A Thesis

to be submitted by

Nagaraj Adiga

Roll. No.: 11610235

for the award of the degree of

DOCTOR OF PHILOSOPHY

January 9, 2017

DEPARTMENT OF ELECTRONICS AND ELECTRICAL ENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY GUWAHATI

GUWAHATI - 781 039

1 Introduction

This thesis proposes the glottal activity region based processing for speech synthesis. The glottal activity regions are perceptually significant and comprise majority of speech sounds. Hence, distinct features derived from the glottal activity regions are used for speech synthesis. In particular, the thesis is focused on improving the voice quality of statistical parametric speech synthesis (SPSS) in glottal activity region. In SPSS, hidden Markov model (HMM) based statistical model is commonly used [1]. The naturalness of speech is mainly attributed by the source signal. In a conventional way, it is modeled by the impulse excitation [1, 2]. It represents only one aspect of the source signal, namely periodic component, however, source signal consist of other attributes like aperiodic component and phase component [3–5]. It may be difficult to capture all these aspects using the conventional features based on segmental processing of speech, which captures the average information of speech production system. The intelligibility of speech i.e message information is represented by the vocal tract system. The features used to represent vocal tract systems like linear prediction coefficients and Mel-cepstral coefficients are processed by segment wise and may not capture the coarticulation effect of production mechanism efficiently [1]. In addition, the excitation design of SPSS is currently based on a two-state model which depends on the accuracy of the voicing decision and the fundamental frequency [6]. The failures like, voiced region classifying as an unvoiced region gives hoarseness to voice quality, whereas unvoiced region classifying as voiced region gives buzziness. Therefore, there need to be suprasegmental features which do voicing decision and fundamental frequency accurately.

2 Motivation for glottal activity region based processing

Speech synthesis using SPSS is inevitable due to the generalized framework and its capability for adaptation of models [1, 2]. However, shortcomings of SPSS are with the issues like vocoded synthetic quality due to poor source-filter features, over-smoothing of parameters in the SPSS, and simplified vocoder model [1, 2]. Based on the results of STRAIGHT and GLOTTHMM vocoder work [7, 8] naturalness and intelligibility of SPSS can be improved by employing better source-filter features and advanced vocoder framework. Therefore, in this thesis emphasis is on the better features to represent the source-filter modules to get high voice quality. For addressing these issues mentioned in SPSS, the motivation for the present work in this thesis are the following:

- In general most state-of-the-art algorithms uses voicing decision based on the pitch or voicing strength computed over a segment of 20 to 30 ms [9]. However, these parameters of voiced sounds vary even for each glottal cycle and may present for shorter duration particularly in voiced consonants. Hence, to characterize these different aspects of glottal activity region, sub-segmental analysis is to be done. In addition, voiced sounds characterized by different aspects of glottal activity like duty cycle (open quotient) and skewness (speed quotient), which can be incorporated along with existing parameters [4, 5].

- In SPSS, vocal tract is modeled by Mel-cepstral coefficients (MCEP) computed from short-time Fourier transform (STFT) spectrogram, which still contains glottal source information due to fixed segmental analysis. In addition, STFT spectrogram does not capture the co-articulation mechanism effectively, which may result in lower intelligibility [7]. Hence, an alternative method is required to capture coarticulation aspects of speech production mechanism, which completely deemphasize the effect of the glottal source signal in vocal tract spectrum.

- The naturalness or speaker characteristics are mainly attributed by the glottal source modeling [8]. The naturalness of conventional SPSS is lagging behind concatenative speech synthesis mainly due to simple impulse/noise source model [10]. In addition, conventional glottal source model ignores different aspects of glottal activity events like aperiodic component and phase component. There needs to be a mechanism to compactly preserve the glottal source signal or an alternate mechanism to preserve the aperiodic component and phase component to improve the naturalness.

Motivated by these observations, the primary objective of this thesis is to demonstrate the significance of processing glottal activity regions to derive suprasegmental, system, and source features to depict the speech production mechanism in a better way and in turn improve prosody, naturalness, and intelligibility of SPSS framework.

3 Contributions of the Thesis

The major contributions of this thesis are as follows:

- Glottal activity region detection using three glottal source features, namely, the strength of excitation (SoE), normalized autocorrelation peak strength (NAPS), and higher-order statistics (HOS).

- Using glottal activity region detection as a voicing indicator and improving the accuracy of voicing decision with the classifiers. Finally, applying voicing decision for speech synthesis in an SPSS framework.

- 2-D based processing of speech spectrogram using Riesz transform to get the smoothed vocal tract envelope. In addition, Riesz transform provides 2-D pitch map, voicing decision, and aperiodicity spectrum. Finally, modeling these Riesz parameters in SPSS and showing its importance for improving the quality of SPSS.

- Source modeling by different aspects of glottal activity region like epoch location, epoch strength, aperiodic component, and phase information using integrated linear prediction residual (ILPR). Periodic components are modeled using MCEP and aperiodic representation of ILPR signal in glottal activity region is modeled using white Gaussian noise modulated with the pitch adaptive triangular envelope weighted by SoE. Finally, processing of phase component present in the ILPR using all-pass filter coefficients (APC) and showing its significance to SPSS.

- Combining suprasegmental, source, and system features to improve the prosody, naturalness, and intelligibility of SPSS, respectively.

4 Glottal activity region detection

The first work of the thesis deals with detection of glottal activity regions using glottal source features. The features SoE, NAPS, and HOS are the representation of energy, periodicity, and asymmetrical nature of glottal cycle and their combination is proposed to identify the glottal activity region. To capture the glottal activity features directly from the source signal, the effect of time-varying vocal tract response present in the speech signal is to be minimized. Hence, zero-frequency filter (ZFF) and inverse linear prediction (LP) filter are applied independently on speech to get two different source representations, zero-frequency filtered signal (ZFFS) and ILPR signal [11–13], respectively. The features SoE and NAPS are applied on ZFFS, whereas, HOS is applied on the ILPR signal. The combination of all three features is shown for glottal activity region task on both differentiated electroglottograph (DEGG) and the speech signal. The results of the proposed method in comparison to the other state-of-the-art algorithms are shown in Table 1 [14–16]. The proposed combination of glottal source features performed better both in DEGG and speech signal when compared to other methods.

Table 1: Comparison of the proposed method with other methods in clean conditions represented in terms of GAD frame error (in %)

Database	Signal	wavesurfer	SRH	REAPER	Proposed
CMU-ARCTIC Male	DEGG	8.21	6.86	4.93	1.74
	speech	8.05	6.36	5.14	3.47
CMU-ARCTIC Female	DEGG	4.69	3.01	2.62	1.76
	speech	6.49	4.17	3.53	2.72
PTDB-TUG Male	DEGG	11.80	12.18	13.62	2.72
	speech	12.08	12.09	6.14	4.21
PTDB-TUG Female	DEGG	8.11	6.48	5.05	2.47
	speech	6.92	6.52	5.08	4.27
Average	DEGG	8.20	7.13	6.55	2.17
	speech	8.38	7.28	4.97	3.66

5 Glottal activity features for speech synthesis

The different glottal activity features like epoch locations, epoch strength, phase information, and voicing decision are useful for high quality of speech signal. In existing methods, voicing detection relies mostly on fundamental frequency F_0, which may result in errors when the prediction is inaccurate [17]. The voicing decision is computed from the different glottal activity features present in the excitation source signal. The glottal activity features SoE, NAPS, and HOS are used for the voicing decision. To improve the voicing decision and to avoid the heuristic threshold for classification, glottal activity features are trained using different statistical learning methods such as a k-nearest neighbor, support vector machine (SVM), and deep belief network [18–20]. The voicing detection performs best with SVM classifier and its effectiveness is tested by using it as voicing decision for SPSS. The glottal activity features SoE, NAPS, and HOS are modeled in HMM along with F_0 and MCEP to get the voicing decision. Here, the glottal activity features are modeled using continuous distribution instead of conventional multi-space distribution F_0 model [6,17]. The objective and subjective evaluations demonstrate that the proposed method improves the naturalness of synthetic speech. Figure 1 shows the mean opinion score (MOS) in comparison to other stat-of-the-art algorithms [9,15–17]. The figure shows that the proposed method performed better both in male and female speakers.

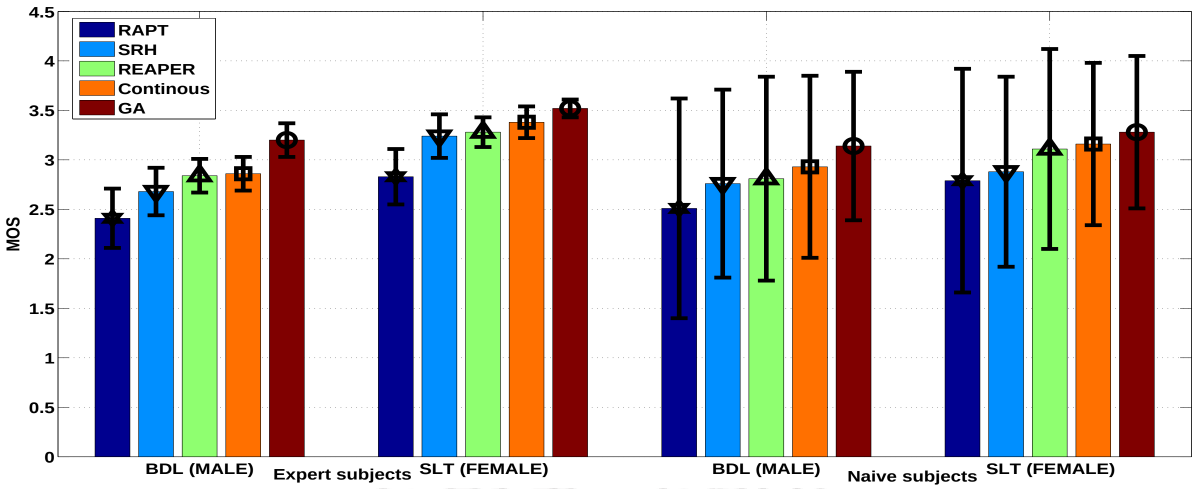

Figure 1: Average MOS of five SPSS systems with RAPT, STRAIGHT, REAPER, Continuous, and GAD model based voicing decision, respectively, for SLT (female speaker) and BDL (male speaker)

6 Riesz Transform for Speech synthesis

The traditional analysis by synthesis methods are based on the fixed short time frame analysis, resulting errors in formant estimation. Hence, 2-D spectro-temporal analysis by synthesis method is proposed using Riesz transform [21,22]. The 2-D spectro-temporal analysis is motivated by the fact that the human auditory cortex is tuned to localized spectro-temporal modulations. The spectro-temporal receptive fields of these cortical cells look like 2-D spectro-temporal Gabor filters [23]. The demodulation of 2-D spectro-temporal patches using Riesz transform separates the carrier signal from the speech spectrum present in the glottal activity region. Hence, yields smoothed spectral envelope, carrier signal, and coherence map, representing vocal tract spectrum, source signal, and periodicity, respectively, using a single framework.

To model the vocal tract information in the SPSS, compact representation of spectral envelope into few parameters is necessary. Hence, MCEP are calculated from spectral envelope extracted using the Riesz transform [24]. The spectrogram is computed with 20 ms Hamming window with a shift of 1 ms and it is demodulated using Riesz transform to obtain smoothed spectral envelope free of carrier component due to glottal activity. Figure 2 shows the spectrogram computed for a sentence taken from ARCTIC database [25], which is showing the smoothed extracted AM envelope and the harmonic structure of carrier in the glottal activity region. The obtained carrier signal is used for estimation of voicing decision, F_0, and aperiodicity estimation. The aperiodicity spectrum is parametrized using in the glottal activity region with 25 MCEP since in the non-glottal region carrier signal is absent. The parametrized envelope MCEP and aperiodicity MCEP are modeled in SPSS.

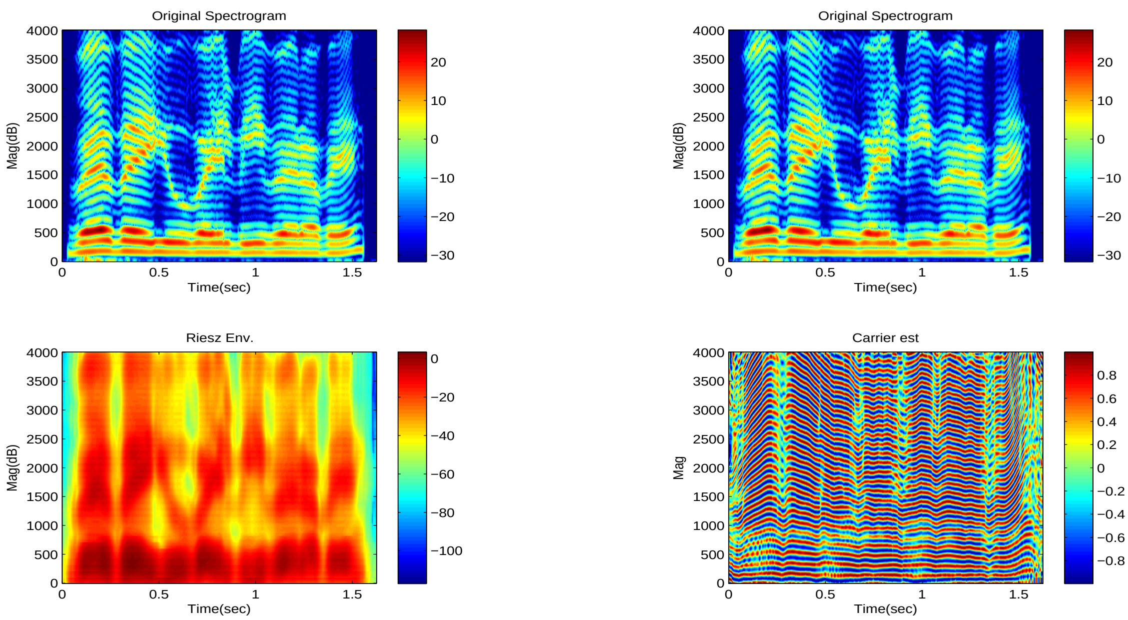

Figure 2: The spectrogram demodulation to estimate smooth AM and harmonic carrier using Riesz transform

The PESQ and SNR scores of synthesized files for proposed Riesz method and STRAIGHT method are given in Table 2. The overall quality from the aperiodic excitation with Riesz envelope is better than STRAIGHT method for the male speaker. The STRAIGHT method is giving slightly better perceptual quality for the female speaker. Further, the results show that for voiced sounds, proposed method performed better than STRAIGHT system. Overall for continuous speech proposed method performed equally well. The effectiveness of Riesz transform

Table 2: PESQ and SNR scores of Riesz and STRAIGHT methods for Indian TTS database

Objective Measure	**Riesz**		**STRAIGHT**	
	Male	**Female**	**Male**	**Female**
PESQ	3.78	3.40	3.47	3.57
SNR (dB)	16.23	14.76	13.23	14.41

is further studied in statistical framework using HMM-based speech synthesis and results show that decomposition of spectrum into an envelope, carrier, a periodic map is equally effective like STRAIGHT method.

7 Integrated linear prediction residual for source modeling

Source modeling for SPSS is proposed using ILPR signal. The nature of ILPR waveform resembles the glottal flow derivative signal and may keep the speaker characteristics in a better way [13]. The different events are present in the source signal, namely, glottal closure, glottal opening, onset of burst, frication and a small number of excitation instants around them [26]. The speech signal is processed independently by ZFF to obtain different events present in speech production. These events are used as anchor points for extracting the different aspects of speech production like strength around epochs and aperiodic component generation. The ILPR signal is modeled in the frequency domain by dividing the spectrum into two bands to characterize periodic and aperiodic components of the voiced speech segment. The periodic components of ILPR signal below the maximum voiced frequency (f_m) are modeled using MCEP, whereas aperiodic component above f_m is modeled by pitch adaptive triangular noise envelope weighted by the SoE. The MCEP and SoE of residual signal are modeled on the HMM framework along with MCEP and F_0 representing vocal tract information and fundamental frequency, respectively. The mean opinion score (MOS) of the proposed method is shown in Figure 3. The synthesized speech by the proposed source modeling reduces the buzziness and improves the speaker similarity compared to the conventional impulse/noise and mixed excitation source modeling and comparable with STRAIGHT based excitation [27].

The reason for the better MOS in STRAIGHT method is due to the fact that phase component is also integrated into excitation design, whereas the proposed method phase component is not modeled. Hence, the cosine phase of ILPR signal is modeled using the all-pass filter coefficients, by assuming the it as the output of all-pass filter. Here, both the input and all-pass filter coefficients (APC) are unknown. Hence, the energy of input signals is expressed in terms of objective function, which is an internally function of APC. To minimize entropy, the gradient descent algorithm is used and finally APC are optimized to get the minimum error [28]. The APC are modeled along with MCEP coefficients in SPSS. The MOS and preference test (PT) of the proposed method are shown in Table 3. Both the subjective tests show that cosine phase

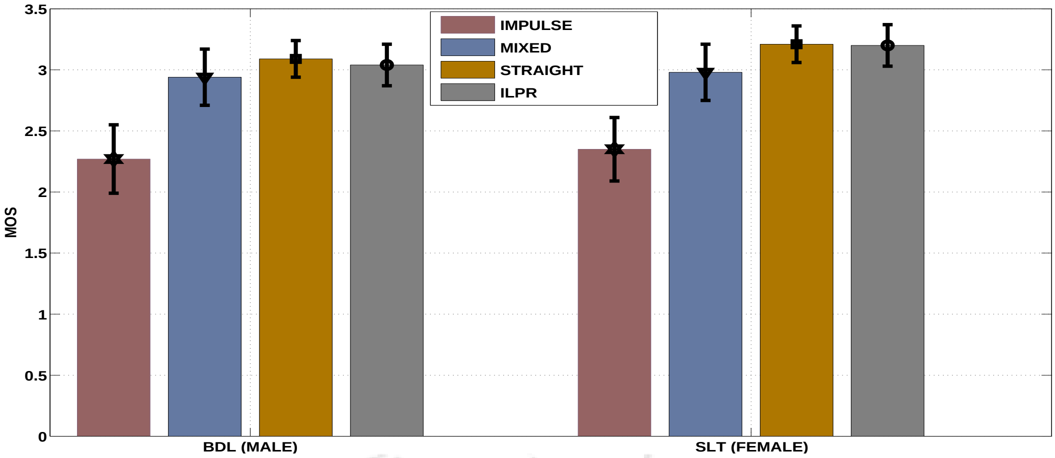

Figure 3: Average MOS of 4 SPSS systems, impulse/noise, mixed, STRAIGHT and ILPR,respectively,for SLT and BDL speaker

model is better than the STRAIGHT method.

Table 3: Subjective evaluation results of MOS and PT for different SPSS systems with 95% confidence interval

Experimental Evaluation	**Different types of phase information**				**p value**
	Impulse	**STRAIGHT**	**Cosine**	**Same**	
MOS	2.71±1.01	3.15±0.92	3.28±0.81	-	
PT	11%	-	80%	9%	2.16×10^{-9}
	26%	64%	-	10%	9.22×10^{-7}
	-	39%	48%	13%	3.67×10^{-2}

8 Combination of Suprasegmental, system, and source features for Speech synthesis:

A combined framework of speech synthesis is demonstrated by processing speech in glottal activity region. Glottal activity regions constitute the majority of speech sound units and these regions are perceptually very important for high voice quality. The glottal activity features are broadly categorized as suprasegmental, system, and source features, which essentially represent the prosodic, intelligibility, and naturalness of speech, respectively. By combining the various glottal activity features aids in bringing the advantages present in each of these features to the synthesis system and getting best features results in the enhancement of overall perceptual quality of SPSS.

The analysis and synthesis block diagram of proposed glottal activity region based speech synthesis framework is shown in Figure 4 and 5, respectively. The objective results of the two systems using the proposed and STRAIGHT framework are presented in Table 4. It is observed that proposed method performs better than STRAIGHT method in terms of lesser

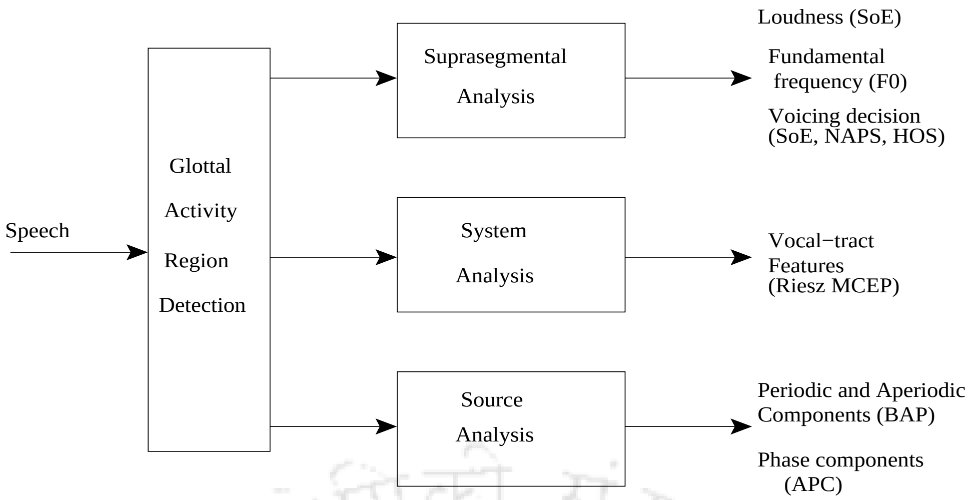

Figure 4: Proposed Analysis framework for Glottal activity based processing for SPSS

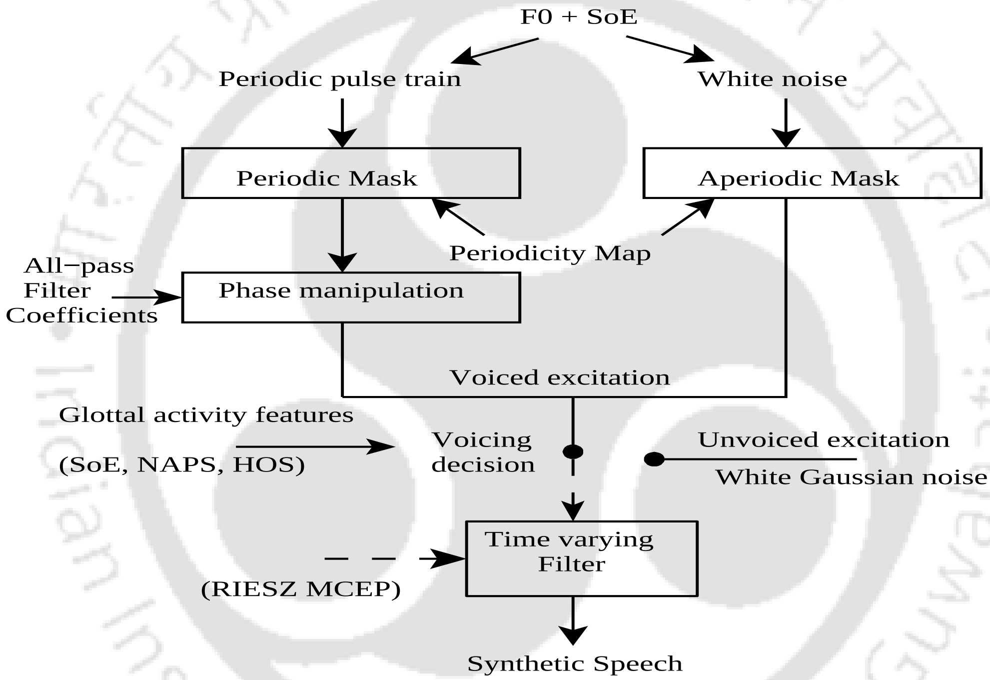

Figure 5: Proposed Synthesis framework for Glottal activity based processing for SPSS

Mel-cepstral distance (MCD), F_0 root mean square error (RMSE), and voicing error (V/UV). In the case of aperiodicity parameter, STRAIGHT method gives lesser band aperiodicity (BAP) spectrum error. In general, objective results confirms that the proposed method is better than STRAIGHT method.

Table 4: Comparison of Objective results using STRAIGHT and proposed analysis/synthesis framework for SPSS

Objective Measure	**MCD (dB)**	**BAP (dB)**	F_0 **RMSE Hz**	**V/UV %**
STRAIGHT	4.36	1.40	9.31	11.66
Proposed method	4.12	2.05	8.55	4.18

9 List of Publications

- Journals Published and communicated

  1 N. Adiga and S. R. M. Prasanna, "Detection of glottal activity using different attributes of source information," IEEE Signal Processing Letters, vol. 22, no. 11, pp. 2107-2111, Nov 2015

  2 N. Adiga, B. K. Khonglah and S. R. M. Prasanna, "Improved voicing decision using glottal activity features for HMM based speech synthesis" under first revision in *Journal of Digital Signal Processing.*

  3 N.Adiga and S. R. M. Prasanna, "Acoustic Features for Statistical Parametric Speech Synthesis: A Review" under first revision in *IETE Technical review.*

- Conferences

  1 N. Adiga and S. R. M. Prasanna, "Significance of instants of significant excitation for source modeling," in Proc. INTERSPEECH, 2013.

  2 N. Adiga and S. R. M. Prasanna, "Epochs Based Compression of LP Residual for Source Modeling in Text-to-Speech Synthesis", in Proc NCC, 2014.

  3 N. Adiga, D.Govind and S. R. M. Prasanna, "Significance of Epoch Identification Accuracy for Prosody Modification", in Proc. SPCOM, 2014.

  4 N. Adiga and S. R. M. Prasanna, "A Hybrid Text-to-Speech Synthesis using Vowel and Non Vowel like regions", in Proc. INDICON, 2014.

  5 N. Adiga and S. R. M. Prasanna, "Source modeling for HMM based speech synthesis using Integrated LP Residual", in Proc. ICASSP, 2016.

References

[1] K. Tokuda, Y. Nankaku, T. Toda, H. Zen, J. Yamagishi, and K. Oura, "Speech synthesis based on hidden Markov models," *Proc. IEEE*, vol. 101-5, pp. 1234–1252, 2013.

[2] H. Zen, K. Tokuda, and A. W. Black, "Statistical parametric speech synthesis," *Speech Commun.*, vol. 51(11), pp. 1039–1064, 2009.

[3] A. Krishnamurthy and D. Childers, "Two-channel speech analysis," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 34, no. 4, pp. 730–743, Aug 1986.

[4] T. V. Ananthapadmanabha, "Acoustic analysis of voice source dynamics. STL-QPSR 2–3," Speech, Music and Hearing, Royal Institute of Technology, Stockholm, Tech. Rep., 1984.

[5] M. Rothenberg, "Glottal noise during speech," *Quarterly Progress Status ReportSpeech Transmission Laboratory of the Royal Institute of Technology, Stockholm*, p. 1, 1974.

[6] K. Tokuda, T. Masuko, N. Miyazaki, and T. Kobayashi, "Hidden Markov models based on multi-space probability distribution for pitch pattern modeling," in *Proc. Int. Conf. Acoust. Speech Signal Process.*, vol. 1, Mar 1999, pp. 229–232 vol.1.

[7] H. Kawahara, I. Masuda-Katsuse, and A. de Cheveigné, "Restructuring speech representations using a pitch-adaptive time frequency smoothing and an instantaneous-frequency-based F0 extraction," *Speech Commun.*, vol. 27(3-4), pp. 187–207, 1999.

[8] T. Raitio, A. Suni, J. Yamagishi, H. Pulakka, J. Nurminen, M. Vainio, and P. Alku, "HMM-based speech synthesis utilizing glottal inverse filtering," *IEEE Trans. Audio Speech Lang. Process.*, vol. 19-1, pp. 153–165, 2011.

[9] D. Talkin, "A robust algorithm for pitch tracking (RAPT)," *Speech coding and synthesis*, vol. 495, p. 518, 1995.

[10] A. J. Hunt and A. W. Black, "Unit selection in a concatenative speech synthesis system using a large speech database," *Proc. Int. Conf. Acoust. Speech Signal Process.*, vol. 1, pp. 373–376, 1996.

[11] K. S. R. Murthy and B. Yegnanarayana, "Epoch extraction from speech signals," *IEEE Trans. Audio Speech Lang. Process.*, vol. 16, pp. 1602–1613, November 2008.

[12] K. S. R. Murthy, B. Yegnanarayana, and M. A. Joseph, "Characterization of glottal activity from speech signals," *IEEE Signal process. lett.*, vol. 16, no. 6, pp. 469–472, June 2009.

[13] A. Prathosh, T. Ananthapadmanabha, and A. Ramakrishnan, "Epoch extraction based on integrated linear prediction residual using plosion index," *IEEE Trans. Audio Speech Lang. Process.*, vol. 21, no. 12, pp. 2471–2480, Dec 2013.

[14] K. Sjölander and J. Beskow, "Wavesurfer - an open source speech tool," in *Proc. Interspeech*, 2000, pp. 464–467.

[15] T. Drugman and A. Alwan, "Joint robust voicing detection and pitch estimation based on residual harmonics," in *Proc. Interspeech*, 2011, pp. 1973–1976. [Online]. Available: http://www.isca-speech.org/archive/interspeech_2011/i11_1973.html

[16] D. Talkin, "REAPER: Robust epoch and pitch estimator [online] available: https://github.com/google/reaper."

[17] K. Yu and S. Young, "Continuous F0 modeling for HMM based statistical parametric speech synthesis," *IEEE Trans. Audio Speech Lang. Process.*, vol. 19, no. 5, pp. 1071–1079, July 2011.

[18] K. Fukunaga and P. M. Narendra, "A branch and bound algorithm for computing k-nearest neighbors," *IEEE Trans. Computers*, vol. 100, no. 7, pp. 750–753, 1975.

[19] C. Cortes, M. Mohri, and A. Rostamizadeh, "Two-stage learning kernel algorithms," in *Proc. of the 27th Int. Conf. on Machine Learning*, 2010, pp. 239–246.

[20] X.-L. Zhang and J. Wu, "Deep belief networks based voice activity detection," *IEEE Trans. Audio Speech Lang. Process.*, vol. 21, no. 4, pp. 697–710, April 2013.

[21] T. Wang and T. Quatieri, "Two-dimensional speech-signal modeling," *IEEE Trans. Audio Speech Lang. Process.*, vol. 20, no. 6, pp. 1843–1856, Aug 2012.

[22] H. Aragonda and C. Seelamantula, "Demodulation of narrowband speech spectrograms using the riesz transform," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 23, no. 11, pp. 1824–1834, Nov 2015.

[23] S. Shamma, "On the role of space and time in auditory processing," *Trends in cognitive sciences*, vol. 5, no. 8, pp. 340–348, 2001.

[24] S. Imai, "Cepstral analysis synthesis on the mel frequency scale," *Proc. Int. Conf. Acoust. Speech Signal Process.*, vol. 8, pp. 93–96, 1983.

[25] J. Kominek and A. W. Black, "The CMU ARCTIC speech databases," in *Proc. 5th ISCA Speech Synthesis Workshop*, 2004, pp. 223–224. [Online]. Available: http://festvox.org/cmu_arctic/index.html

[26] R. Smits and B. Yegnanarayana, "Determination of instants of significant excitation in speech using group delay function," *IEEE Trans. Speech Audio Process.*, vol. 3(5), pp. 325–333, 1995.

[27] T. Yoshimura, K. Tokuda, T. Masuko, T. Kobayashi, and T. Kitamura, "Mixed excitation for HMM-based speech synthesis," in *Proc. Eurospeech 2001*, 2001.

[28] K. Vijayan, P. R. Reddy, and K. S. R. Murty, "Significance of analytic phase of speech signals in speaker verification," *Speech Commun.*, pp. –, 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167639316000364

Organization of the Thesis

1. Introduction

    1.1 Overview of Text-to-speech Synthesis
    1.2 Issues in the basic version of SPSS
    1.3 Analysis/Synthesis methods
    1.4 Motivation for the present work
    1.5 Organization of the Thesis

2. Acoustic Features for Statistical Parametric Speech Synthesis: A Review

    2.1 Introduction
    2.2 Acoustic feature modeling using different SPSS techniques
    2.3 Suprasegmental features
    2.4 Vocal tract Spectral model
    2.5 Different source models
    2.6 Other Signal Processing Models
    2.7 Summary and Discussion
    2.8 Organization of the Present Work

3. Glottal activity region detection

    3.1 Motivation for glottal activity region detection
    3.2 Features for characterization of glottal activity
    3.3 Robustness of Features for Characterizing Glottal Activity
    3.4 Glottal activity detection using Different Attributes
    3.5 Summary

4. Glottal activity features for Speech Synthesis

    4.1 Motivation for processing glottal activity region
    4.2 Significance of Glottal activity features for Speech Synthesis
    4.3 Glottal activity features for Voicing decision
    4.4 Improvement in the detection of glottal activity region using Classifiers
    4.5 Glottal activity features for Synthesis
    4.6 Experimental evaluation
    4.7 Summary

5. Riesz transform for Speech synthesis

    5.1 Motivation behind the 2-D processing
    5.2 2-D processing using Riesz transform
    5.3 Riesz transform based demodulation for speech spectrum
    5.4 Carrier spectrogram analysis
    5.5 Synthesis methodology
    5.6 Experimental validation
    5.7 Summary

6. Integrated linear prediction residual for Source modeling

    6.1 Source modeling
    6.2 Different components present in Glottal activity region
    6.3 Periodic and Aperiodic modeling
    6.4 Phase modeling
    6.5 Summary

7. Suprasegmental, System, and Source features for Speech Synthesis

    7.1 Introduction
    7.2 Glottal activity region based processing
    7.3 Proposed analysis/synthesis framework
    7.4 Experimental evaluation
    7.5 Summary and Discussion

8. Summary and Conclusions

    8.1 Summary of the work
    8.2 Contributions of this thesis
    8.3 Directions for future work