INTERSPEECH 2017
August 20–24, 2017, Stockholm, Sweden

Direct Modelling of Magnitude and Phase Spectra for Statistical Parametric Speech Synthesis

Felipe Espic, Cassia Valentini-Botinhao, and Simon King

The Centre for Speech Technology Research (CSTR), University of Edinburgh, UK

felipe.espic@ed.ac.uk, cvbotinh@inf.ed.ac.uk, Simon.King@ed.ac.uk

Abstract

We propose a simple new representation for the FFT spectrum tailored to statistical parametric speech synthesis. It consists of four feature streams that describe magnitude, phase and fundamental frequency using real numbers. The proposed feature extraction method does not attempt to decompose the speech structure (e.g., into source+filter or harmonics+noise). By avoiding the simplifications inherent in decomposition, we can dramatically reduce the "phasiness" and "buzziness" typical of most vocoders. The method uses simple and computationally cheap operations and can operate at a lower frame rate than the 200 frames-per-second typical in many systems. It avoids heuristics and methods requiring approximate or iterative solutions, including phase unwrapping.

Two DNN-based acoustic models were built - from male and female speech data - using the Merlin toolkit. Subjective comparisons were made with a state-of-the-art baseline, using the STRAIGHT vocoder. In all variants tested, and for both male and female voices, the proposed method substantially outperformed the baseline. We provide source code to enable our complete system to be replicated.

Index Terms: speech synthesis, vocoding, speeech features, phase modelling, spectral representation.

1. Introduction

In statistical parametric speech synthesis (SPSS), the vocoder has been identified a cause of "buzziness" and "phasiness" [1].

Most popular vocoders in SPSS are based on a source-filter model. The filter is often realised as minimum phase, derived by cepstral analysis of a smooth spectral envelope [2]. The source could be derived from the residual [3, 4] or glottal signal [5], but is commonly just a pulse train / noise, alternating [6] or mixed [7]. An alternative to the source-filter paradigm is sinusoidal modelling [8, 9, 10, 11] which unfortunately has a time-varying number of parameters. Both poorly model aperiodicity.

The use of decorrelating and dimensionality reducing cepstral analysis arises from requirements of Gaussian models. Recently, the adoption of non-parametric models has opened up possibilities for using higher-dimensional, correlated representations. In [12] a restricted Boltzmann machine (RBM) was used to model the spectral envelope distribution, and the reintroduction of neural networks [13] subsequently lead to work modelling higher dimensional representations [14, 15] or modelling a conventional cepstral representation whilst optimising a cost function in the the waveform domain [16, 17].

In [18] a neural network generates 8-bit quantised waveform samples directly in the time domain, with promising results, but at high computational cost, requiring a large database, and with quantisation noise evident. It is not obvious how to design a perceptually-relevant cost function in the waveform domain: one of many challenges faced by this approach.

Recently, we proposed a new waveform generation method for text-to-speech (TTS) in which synthetic speech is generated by modifying the fundamental frequency and spectral envelope of a natural speech sample to match values predicted by a model [19]. This simple method avoids the typical but unnecessary decomposition of speech (source-filter separation) and requires no explicit aperiodicity modelling. Subjective results indicated that it outperforms the state-of-the-art benchmark [2]. This motivated us to keep looking for ever-simpler methods for waveform generation, in the spirit of end-to-end speech synthesis, but without the challenges of direct waveform generation.

We now propose a method to model speech directly from the discrete Fourier transform. We map the complex-valued Fourier transform to a set of four real-valued components that represent the magnitude and phase spectrum. We make no assumptions about the structure of the signal, except that it is partly periodic. We perform none of the typical decompositions (e.g., source-filter separation, harmonics+noise).

2. Proposed Method

The goals for the proposed method are to:

- minimise the number of signal processing / estimation steps;
- extract consistent features suitable for statistical modelling (e.g., they can be meaningfully averaged);
- eliminate "phasiness" and "buzziness", typical of many vocoders;
- work with any standard real-valued deep learning method.

2.1. Challenges

The first obstacle is that neural networks typically used in SPSS only deal with real-valued data, whilst the FFT spectrum is complex. An exploratory study on complex-valued neural networks for SPSS [20] did not achieve competitive results.

Naively treating the real and imaginary parts of the FFT spectrum as separate (real-valued) feature streams would mean that phase is poorly represented and the cost function for phase during training would be biased towards frames with large magnitudes. So, an explicit representation of phase is required.

The first difficulty in dealing with phase comes from the relative time delay between the signal (e.g., glottal pulses) and analysis frame placement. It is necessary to "normalise" the delay over all measured phase spectra, so that the extracted phase values are comparable. Group delay – Figure 2(c) – can be applied to achieve this, but algorithms to calculate this rely on heuristics and are error prone [21, for example].

Figure 2(a) illustrates why the use of wrapped phase is meaningless. Unwrapping phase relies on heuristics and is notoriously error-prone: Figure 2(b).

http://dx.doi.org/10.21437/Interspeech.2017-1647

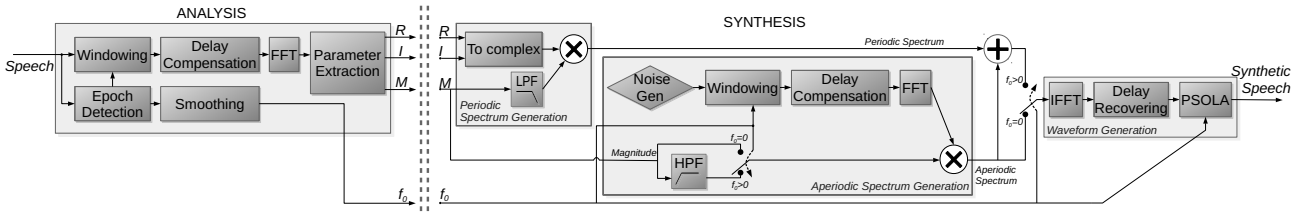

Figure 1: *Diagrams of the analysis and synthesis processes. Four features:* f_0, M, R, *and* I *are extracted to synthesise speech.*

2.2. Analysis

In our proposed method, analysis is pitch-synchronous and results in four feature streams: (1) *Fundamental Frequency,* f_0; (2) *Magnitude Spectrum,* M; (3) *Normalised Real Spectrum,* R; (4) *Normalised Imaginary Spectrum,* I. In the following subsections each is defined, and the complete analysis process depicted in Figure 1 is explained step by step.

2.2.1. Epoch Detection and f_0 Calculation

Analysis frames are centred on epochs in voiced segments, and evenly spaced for unvoiced segments. We use REAPER[1] for epoch detection, although simpler methods could be applied. f_0 is found from the detected epochs by $f_{0[t]} = (e_{[t]} - e_{[t-1]})^{-1}$, where t is the current frame index, f_0 is the fundamental frequency, and e is the epoch location in seconds. Median smoothing with a window of 3 frames length is then applied.

2.2.2. Windowing

Each analysis frame spans two pitch periods.The maximum of a non-symmetrical Hanning window is placed at the central epoch, and its extremes at the previous and next epochs.

The Hanning window does not remove the harmonic structure in the resulting FFT spectrum [22], but it substantially reduces its prominence, thus the FFT spectrum is suitable for acoustic modelling.

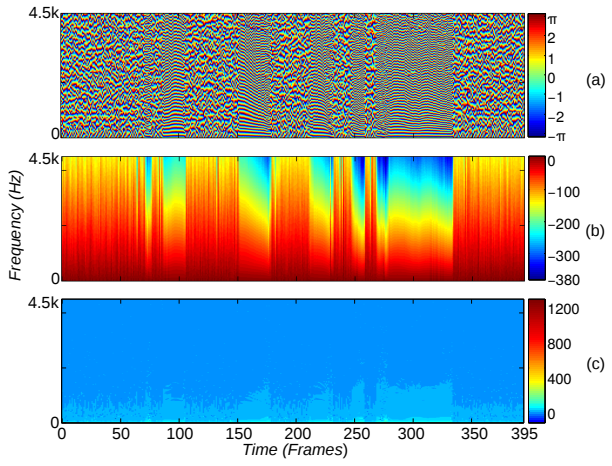

Figure 2: *Examples of typical phase representations extracted from a utterance. The plots show the lack of recognisable patterns that may be successfully used in statistical modelling. (a) Wrapped phase. (b) Unwrapped phase. (c) Group delay.*

[1] https://github.com/google/REAPER

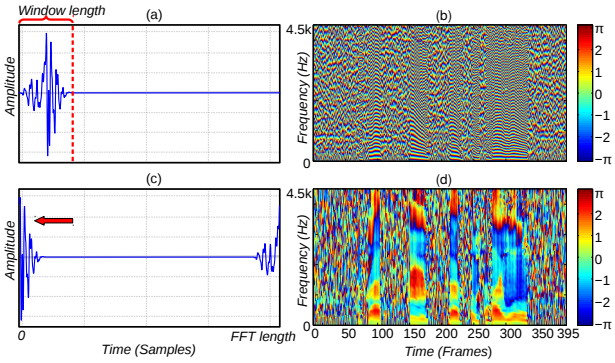

Figure 3: *An example of delay compensation and its effects on the phase spectrogram. Frame before delay compensation in (a) and its wrapped phase spectrogram in (b). Frame after delay compensation in (c) and its phase spectrogram in (d) where clearer phase patterns emerge.*

2.2.3. Delay Compensation

Phase modelling must have inter-frame consistency, so that the phase extracted from different frames can be compared, averaged, etc. Delay has a detrimental effect, so must be normalised.

Group delay compensation is error prone (Section 2.1). The method we use here can be seen as a simple and robust method for group delay compensation. Each windowed frame of the speech signal is zero padded to the FFT length. Then, assuming that epochs are located at points of maximum absolute amplitude within a frame, and treating the frame as a circular buffer, the signal is shifted such that its central epoch is positioned at the start of the frame (See Figure 3). The benefits are:

- phase consistency between frames;
- minimises phase wrapping;
- maximises smoothness of the phase spectrum.

2.2.4. Parameter Extraction

After delay compensation, the **magnitude** spectrum M is computed from the FFT coefficients X in the usual way: $\mathrm{M} = abs(\mathrm{X})$. As noted in Section 2.2.3, consistency matters for **phase** features. We wish to avoid all the typical drawbacks of using wrapped-phase, unwrapped-phase or group delay approaches (Section 2.2.3) and our objectives are:

- consistency – if the phases of two components are close, they need to approach the same numerical value;
- to avoid heuristics.

We start from the wrapped phase obtained from the FFT, which cannot be used directly, since statistical models (e.g.,

DNN) assume consistency. For example, phases near π or $-\pi$ have very different numerical values, even though they are very close in phase domain. To alleviate this, we take the $cosine$ of the phase to map phase values close to π or $-\pi$ to a consistent value around -1 that changes smoothly even when phase jumps (wraps) from π to $-\pi$. This representation is ambiguous regarding sign of the phase, so we add a $sine$ representation. The phase is thus simply described by the normalised real and imaginary parts of the FFT spectrum (complex phase), $\mathrm{R} = Real\{\mathrm{X}\}/abs(\mathrm{X})$, and $\mathrm{I} = Imag\{\mathrm{X}\}/abs(\mathrm{X})$.

Before statistical modelling, magnitude M and fundamental frequency f_0 are transformed to somewhat more perceptually relevant scales by a log transform, as is common in SPSS.

2.3. Synthesis

The synthesis process consists of three main processes: Figure 1. *Periodic spectrum generation* produces the complex spectrum for voiced speech components up to a maximum voiced frequency (MVF). *Aperiodic spectrum generation* uses M and f_0 features, plus phase extracted from random noise to produce the complex spectrum for unvoiced speech, as well as for frequencies above the MVF for voiced speech. Finally, *waveform generation* takes a sequence of complex spectra as input and generates a waveform.

2.3.1. Periodic Spectrum Generation

Complex phase spectrum $\mathrm{P} = (\mathrm{R} + \mathrm{I} \cdot j)/\sqrt{\mathrm{R}^2 + \mathrm{I}^2}$ is computed. The normalisation term in the denominator is needed because P may not be unitary if R and I were generated from a model (e.g., when performing TTS). The predicted magnitude spectrum M is low-pass filtered at the MVF, then multiplied by P (which carries predicted phase), resulting in the complex spectrum for the periodic component. This is only done in voiced segments.

2.3.2. Aperiodic Spectrum Generation

For both voiced and unvoiced segments an aperiodic component is predicted. The phase of aperiodic components is not recovered from R and I features, since it is chaotic and unpredictable. Instead, the aperiodic phase is derived from zero-mean and uniformly distributed random noise. Its dispersion is irrelevant, since its magnitude spectrum will be normalised later. Once generated, this pure zero-mean random noise, is framed and windowed as in analysis (Section 2.2.2).

In voiced segments frames are centred at the epoch locations given by $e_{[t]} = \sum_{i=0}^{t} f_{0[i]}^{-1}$, where $e_{[t]}$ is the epoch location at frame t, and $f_{0[i]}$ is the predicted f_0 at frame i. We observed that in natural voiced speech, the time-domain amplitude envelope of aperiodic components (above the MVF) is pitch synchronous, and energy is concentrated around epoch locations. To emulate this, a narrower window, $w_{[t]} = \left(barlett_{[t]}\right)^{\lambda}$ with $\lambda > 1$ is used. As a consequence, the amplitude envelope of the noise will be shaped during reconstruction accordingly. **In unvoiced segments** frames are uniformly spaced (e.g., 5ms) and windowed with Hanning window.

For both voiced and unvoiced segments the FFT complex spectra of the windowed noise is computed. Then, a spectral mask is constructed to modify its magnitude. Since the noise is generated with an arbitrary dispersion, the average RMS of the magnitude spectrum of the noise is used as a normalising term for the spectral mask. **In voiced segments** the spectral mask is the predicted (by the DNN) magnitude spectrum M, high-pass filtered at the MVF. This filter is complementary to the one used in the *periodic signal generation* stage, and is applied in the same form. **In unvoiced segments** the mask is just the predicted magnitude spectrum M.

Finally, **for both voiced and unvoiced segments**, the complex noise spectrum and the spectral mask are multiplied to produce the complex spectrum of the aperiodic components. **In unvoiced segments** the complex spectra of periodic and aperiodic components are summed. **For unvoiced segments** only the aperiodic component is used.

2.3.3. Waveform Generation

Each complex spectrum is transformed to the time domain by an IFFT and the resulting waveform is shifted forward by a half the FFT length, to revert the time aliasing produced by the *delay compensation* during analysis. The central epoch of the waveform is thus placed centrally in the frame. The final synthetic waveform is constructed by Pitch Synchronous Overlap and Add (PSOLA) driven by the epochs locations obtained from f_0.

3. Experiments

Two neural network-based text-to-speech (TTS) voices were built using the Merlin toolkit [23] from speech data at 48kHz sample rate. A male voice, "Nick" was built by using 2400, 70 and 72 sentences for training, validation, and testing, respectively. A female voice, "Laura" was built with 4500, 60, and 67 sentences. The network architecture was an enhanced version of the simplified long-short term memory (SLSTM) introduced in [24]. We used 4 feedforward layers each of 1024 units, plus an SLSTM recurrent layer of 512 units.

The baseline system operates at a 5ms constant frame rate, with analysis and synthesis performed by STRAIGHT [25, 2], with speech parameters: 60 Mel-cepstral coefficients (MCEPs), 25 band aperiodicities (BAPs), log fundamental frequency (lf_0). This configuration is widely used and is one of the standard recipes included in the Merlin toolkit.

The speech parameters for the proposed system were: FFT-length=4096, aperiodic voiced window factor λ=2.5, MVF=4.5kHz. The spectral features were Mel-warped by transforming the full resolution spectra into MGCs using SPTK[2] (α=0.77) and transforming back to the frequency domain using the fast method described in [14].

The acoustic features for the proposed method were: 60-dimensional log magnitude spectrum (evenly spaced on a Mel scale from 0Hz to Nyquist), 45-dimensional normalised real spectrum (0Hz to MVF), 45-dimensional normalised imaginary spectra (0Hz to MVF), and lf_0. All spectral features are Mel-warped. The normalised real and imaginary spectra are zero-interpolated for unvoiced speech (done during analysis).

The proposed method works pitch synchronously. For the male speaker, this decreases the average number of frames per second by 31.5% compared to the baseline. For the female, both systems are comparable.

3.1. Subjective Evaluation

A subjective evaluation was carried out to measure the *naturalness* achieved by several configurations of the proposed method, and the state-of-the-art baseline.

[2] `https://sourceforge.net/projects/sp-tk/`

Thirty native English speakers (University students) were recruited to take a MUSHRA-like[3] test. Each subject evaluated 18 randomly selected sentences from the "Nick" and "Laura" test sets, respectively, resulting in 36 MUSHRA screens per subject. The stimuli evaluated in each screen were:

- Nat: **Nat**ural speech (the hidden reference).
- Base: The **Base**line system running at constant frame rate and using STRAIGHT for analysis/synthesis.
- PM: The **P**roposed **M**ethod with settings as described in this paper.
- PMVNAp: The **P**roposed **M**ethod with **V**oiced segments having **N**o **Ap**eriodic component
- PMVNApW: The **P**roposed **M**ethod with **V**oiced segments having **N**o **Ap**eriodicity **W**indow – i.e., without using the narrower analysis window of Section 2.3.2.

For all systems, the standard postfilter included in Merlin was applied [26]. For the male speaker, the postfilter moderately affected unvoiced speech regions, thus high frequencies were slightly boosted to compensate. All synthesised signals were high-pass filtered to protect against spurious components that could appear below the voice frequency range. Audio samples are provided at `http://www.felipeespic.com/demo_fft_feats_IS17`.

3.2. Results

One subject was rejected from the analysis, due to inconsistent scores (Natural<20%). To test statistical significance, the Wilcoxon Signed Rank test with $p<0.05$ was applied. Holm-Bonferroni correction was used because of the large number of tests (18×29 per voice). A summary of the scores is in Table 1. Figure 4 plots the mean, median, and dispersion of the scores per system under test, for each voice.

Table 1: *Average MUSHRA Score Per System in Evaluation*

	System				
Speaker	Nat	Base	PM	PMVNAp	PMVNApW
Male	100	43.6	**51.4**	45.6	49.4
Female	100	32.6	**43.8**	34.6	43.1

Significance tests indicate that all configurations of the proposed method significantly outperform the baseline for both voices. The highest scores were achieved by PM, which was significantly preferred over all other systems, except for the female voice where PM and PMVNApW were not significantly different. PMVNApW was significantly preferred over PMVNAp and the baseline for both voices.

4. Conclusion

We propose a new waveform analysis/synthesis method for SPSS, which encodes speech into four feature streams. It does not require estimation of high-level parameters such as: spectral envelope, aperiodicities, harmonics, etc., used by vocoders that attempt to decompose the speech structure.

It does not require any iterative or estimation process beyond the epoch detection performed during analysis. Indeed, it uses fast operations such as: FFT, OLA, and IFFT.

[3]Code from `http://dx.doi.org/10.7488/ds/1316`

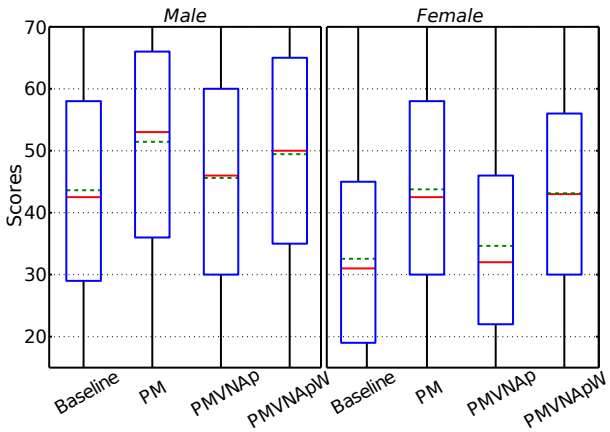

Figure 4: *Absolute scores for the male and female voices. The green dotted line is the mean, and the continuous red line is the median. Natural speech is omitted (mean score is 100) and the vertical scale is limited to 15-70, for clarity.*

Subjective results show the proposed method outperforming a state-of-the-art SPSS system that uses the STRAIGHT vocoder, for a female and a male voice. It largely eliminates "buzziness" and "phasiness", delivering a more natural sound.

The proposed method does not use heuristics or unstable algorithms that are required for methods relying on unwrapped phase or group delay. We demonstrated the importance of proper modelling of aperiodic components during voiced speech, by including in the subjective evaluation variants of our method that did not include this (although they still outperformed the baseline).

In addition, the proposed method decreases the frame rate for any speaker with mean f_0 below 200Hz, with an impressive reduction of 31.5% for our male speaker.

The proposed method, as a reliable representation of the FFT spectrum, might be useful for other audio signal processing applications.

In the future, we plan to extend this work by:

- Eliminating the need for f_0 modelling and prediction.
- Avoiding voiced/unvoiced decisions.
- Avoiding the assumption of a maximum voiced frequency (MVF).

4.1. Reproducibility

A Merlin recipe and additional code that replicates the signal processing and DNN systems presented, are available at `http://www.felipeespic.com/fft_feats_IS17`.

5. Acknowledgements

Felipe Espic is funded by the Chilean National Agency of Technology and Scientific Research (CONICYT) - Becas Chile 72150507.

6. References

[1] T. Merritt, J. Latorre, and S. King, "Attributing modelling errors in HMM synthesis by stepping gradually from natural to modelled speech," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Brisbane, Apr. 2015, pp. 4220–4224.

[2] H. Kawahara, I. Masuda-Katsuse, and A. de Cheveign, "Restructuring speech representations using a pitch-adaptive timefrequency smoothing and an instantaneous-frequency-based F0 extraction: Possible role of a repetitive structure in sounds," *Speech Communication*, vol. 27, no. 34, pp. 187 – 207, 1999.

[3] R. Maia, T. Toda, H. Zen, Y. Nankaku, and K. Tokuda, "An excitation model for HMM-based speech synthesis based on residual modeling," in *Proc. SSW*, Bonn , Germany, August 2007, pp. 131–136.

[4] T. Drugman, A. Moinet, T. Dutoit, and G. Wilfart, "Using a pitch-synchronous residual codebook for hybrid HMM/frame selection speech synthesis," in *Proc. ICASSP*, Taipei, Taiwan, April 2009, pp. 3793–3796.

[5] 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. on Audio, Speech and Language Processing*, vol. 19, no. 1, pp. 153–165, Jan. 2011.

[6] T. Yoshimura, K. Tokuda, T. Masuko, T. Kobayashi, and T. Kitamura, "Simultaneous modeling of spectrum, pitch and duration in HMM-based speech synthesis," in *Proc. Eurospeech*, Budapest, Hungary, September 1999, pp. 2347–2350.

[7] T. Yoshimura, K. Tokuda, T. Masukom, T. Kobayashi, and T. Kitamura, "Mixed excitation for HMM-based speech synthesis," in *Proc. Eurospeech*, Aalborg, Denmark, September 2001, pp. 2263–2267.

[8] C. Hemptinne, *Integration of the Harmonic plus Noise Model (HNM) into the Hidden Markov Model-Based Speech Synthesis System (HTS)*. Martigny, Switzerland: MSc dissertation - IDIAP Research Institute, 2006.

[9] E. Banos, D. Erro, A. Bonafonte, and A. Moreno, "Flexible harmonic/stochastic modeling for HMM-based speech synthesis," in *In V Jornadas en Tecnologias del Habla*, Bilbao, Spain, November 2008, pp. 145–148.

[10] Q. Hu, Y. Stylianou, R. Maia, K. Richmond, J. Yamagishi, and J. Latorre, "An investigation of the application of dynamic sinusoidal models to statistical parametric speech synthesis," in *Proc. Interspeech*, Singapore, September 2014, pp. 780–784.

[11] Q. Hu, K. Richmond, J. Yamagishi, and J. Latorre, "An experimental comparison of multiple vocoder types," in *8th ISCA Workshop on Speech Synthesis*, Barcelona, Spain, August 2013, pp. 155–160.

[12] Z. H. Ling, L. Deng, and D. Yu, "Modeling spectral envelopes using restricted boltzmann machines and deep belief networks for statistical parametric speech synthesis," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 21, no. 10, pp. 2129–2139, Oct 2013.

[13] H. Zen, A. Senior, and M. Schuster, "Statistical parametric speech synthesis using deep neural networks," in *Proc. ICASSP*, Vancouver, BC, Canada, 2013, pp. 7962–7966.

[14] C. Valentini-Botinhao, Z. Wu, and S. King, "Towards minimum perceptual error training for DNN-based speech synthesis," in *Proc. Interspeech*, Dresden, Germany, September 2015.

[15] S. Takaki and J. Yamagishi, "A deep auto-encoder based low-dimensional feature extraction from fft spectral envelopes for statistical parametric speech synthesis," in *ICASSP*, March 2016, pp. 5535–5539.

[16] K. Tokuda and H. Zen, "Directly modeling speech waveforms by neural networks for statistical parametric speech synthesis," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2015, pp. 4215–4219.

[17] ——, "Directly modeling voiced and unvoiced components in speech waveforms by neural networks," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2016, pp. 5640–5644.

[18] A. van den Oord, S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, N. Kalchbrenner, A. W. Senior, and K. Kavukcuoglu, "Wavenet: A generative model for raw audio," *CoRR*, vol. abs/1609.03499, 2016. [Online]. Available: http://arxiv.org/abs/1609.03499

[19] F. Espic, C. Valentini-Botinhao, Z. Wu, and S. King, "Waveform generation based on signal reshaping for statistical parametric speech synthesis," in *Proc. Interspeech*, San Francisco, CA, USA, September 2016, pp. 2263–2267.

[20] Q. Hu, J. Yamagishi, K. Richmond, K. Subramanian, and Y. Stylianou, "Initial investigation of speech synthesis based on complex-valued neural networks," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, March 2016, pp. 5630–5634.

[21] H. A. Murthy and V. Gadde, "The modified group delay function and its application to phoneme recognition," in *Acoustics, Speech, and Signal Processing, 2003. Proceedings. (ICASSP '03). 2003 IEEE International Conference on*, vol. 1, April 2003, pp. I–68–71 vol.1.

[22] D. Erro, I. Sainz, E. Navas, and I. Hernaez, "Efficient spectral envelope estimation from harmonic speech signals," *Electronics Letters*, vol. 48, no. 16, pp. 1019–1021, August 2012.

[23] Z. Wu, O. Watts, and S. King, "Merlin: An open source neural network speech synthesis system," in *Proc. 9th ISCA Speech Synthesis Workshop (SSW9)*, Sunnyvale, CA, USA, September 2016.

[24] Z. Wu and S. King, "Investigating gated recurrent networks for speech synthesis," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2016, pp. 5140–5144.

[25] H. Kawahara, H. Katayose, A. de Cheveigné, and R. D. Patterson, "Fixed point analysis of frequency to instantaneous frequency mapping for accurate estimation of F0 and periodicity," in *Sixth European Conference on Speech Communication and Technology, EUROSPEECH 1999, Budapest, Hungary, September 5-9, 1999*, 1999.

[26] K. Koishida, K. Tokuda, T. Kobayashi, and S. Imai, "Celp coding based on mel-cepstral analysis," in *1995 International Conference on Acoustics, Speech, and Signal Processing*, vol. 1, May 1995, pp. 33–36 vol.1.