

CHARACTERIZATION OF SUBHARMONIC VOICES USING PHASE DERIVATIVES

Hideki Kawahara¹, Ken-Ichi Sakakibara²

¹Wakayama University, Japan

²Health University of Hokkaido, Japan

kawahara@sys.wakayama-u.ac.jp, kis@hoku-iryo-u.ac.jp

Abstract

We propose two measures for characterizing aperiodic phenomena in subharmonic voices using phase derivatives. Phase derivatives have clearer physical meaning than phase itself. The time derivative of the phase is the instantaneous frequency and the frequency derivative of the phase, with negating polarity, is the group delay. Power spectrum weighted average of each representation provides centroid of energy in the frequency or time domain [1,2]. For instantaneous frequency-based measure, we introduce a simple and linear (from 0 dB to 80 dB) SNR estimator based on a six-term cosine series windowing function, which has the maximum side-lobe level of a -14.1 dB and 54 dB/oct decay rate of side-lobes [3]. For group delay-based measure, we introduce a hierarchical set of multi-resolution event detectors using minimum phase time compensation. These measures are extracted based on an initial periodicity map, which is derived using YANG vocoder [4]. These measures were tested using a time-varying speech production simulator, which is a part of an open source education/research tool for speech science, SparkNG [5]. The unique feature of the simulator is its anti-aliased glottal source models [6], which comprises of the L-F model (Fant and Liljencrants) and the F-L model (Fujisaki and Ljungqvist). Background noise level caused by aliasing is lower than -170 dB around the fundamental component [3]. The fundamental frequency trajectory of the simulator can be controlled at the audio sampling rate. The time parameters of the glottal source model can be set for each glottal cycle.

[1] B. Boashash, "Estimating and interpreting the instantaneous frequency of a signal. I. Fundamentals," *Proceedings of the IEEE*, vol. 80, no. 4, pp. 520–538, 1992.

[2] B. Boashash, "Estimating and Interpreting the Instantaneous Frequency of a Signal - Part 2: Algorithms and Applications," *Proceedings of the IEEE*, vol. 80, no. 4, pp. 540–568, 1992.

[3] H. Kawahara, K. Sakakibara, M. Morise, H. Banno, T. Toda, T. Irino, "A new cosine series antialiasing function and its application to aliasing-free glottal source models for speech and singing synthesis," *Interspeech 2017*, [submitted]

[4] H. Kawahara, Y. Agiomyrgiannakis, and H. Zen, "Using instantaneous frequency and aperiodicity detection to estimate F0 for high-quality speech synthesis," *arXiv preprint arXiv:1605.07809*, 2016. [Online]. Available: <http://arxiv.org/abs/1605.07809>

[5] H. Kawahara, H. SparkNG: Matlab realtime speech tools and voice production tools, <http://www.wakayama-u.ac.jp/%7ekawahara/SparkNG/>

[6] H. Kawahara, K. Sakakibara, M. Morise, H. Banno, T. Toda, "A new cosine series antialiasing function and its application to aliasing-free glottal source models for speech and singing synthesis," *Interspeech 2017*, [submitted]