Aliasing reduction in L-F model implementation for an interactive tool applicable to speech science education

Hideki KAWAHARA†, Ken-Ichi SAKAKIBARA‡‡, Hideki BANNO‡‡‡, Masanori MORISE‡‡‡‡, Tomoki TODA‡‡‡‡‡, and Toshio IRINO†

† Wakayama University
‡‡ Health Sciences University of Hokkaido
‡‡‡ Meijo University
‡‡‡‡ University of Yamanashi
‡‡‡‡‡ Nara Institute of Science and Technology

IEICE Technical meeting, Sendai Japan, 3 August 2015
Overview

• Interactive environment for speech science education

• Relation between vocal tract shape, transfer function, pole locations, LSF, F0, vibrato, voice source and speech sounds.

• A closed-form representation of the anti-aliased version of the well known voice source model (L-F model) is introduced

• Matlab-based implementation
Overview

- Interactive environment for speech science education

- Relation between vocal tract shape, transfer function, pole locations, LSF, F0, vibrato, voice source and speech sounds.

- A closed-form representation of the anti-aliased version of the well known voice source model (L-F model) is introduced

- Matlab-based implementation
vocal tract area function (log-area)

mouth

3D shape
actual modification: scaling \times \text{modifier shape}
vocal tract transfer function

poles of the transfer function

selected pole

pole information
LSF: line spectrum frequencies
Overview

• Interactive environment for speech science education

• Relation between vocal tract shape, transfer function, pole locations, LSF, F0, vibrato, voice source and speech sounds.

• A closed-form representation of the anti-aliased version of the well known voice source model (L-F model) is introduced

• Matlab-based implementation

論文

線形予測符号化と複合正弦波モデル化の対称性

嵯峨山茂樹†a） 板倉 文忠††

論文

線形予測符号化と複合正弦波モデル化の対称性

嵯峨山 茂樹†

板倉文 忠††

Voice source model

piecewise (exponential) function: discontinuity

A four-parameter model of glottal flow

Fant, G. and Liljencrants, J. and Lin, Q.

journal: STL-QPSR
volume: 26
number: 4
year: 1985
pages: 001-013

http://www.speech.kth.se/qpsr
Voice source model

piecewise (exponential) function: singularity

Modeling the glottal volume-velocity waveform for three voice types

D. G. Childers
405 CSE, Department of Electrical Engineering, University of Florida, Gainesville, Florida 32611-2024

Chieteuok Ahn
Electronics Telecommunications Research Institute, Video Communications, P. O. Box 8, Daeduk Science Town, Daejeon, 305-606 Korea

(Received 18 January 1994; accepted for publication 17 August 1994)
Voice source model

piecewise (exponential) function: singularity

\[ U_g(t) \]

\[ U_0 \]

\[ \text{Slope} = -E_e \]

\[ T_d = U_0/E_e \]

\[ U_g'(t) \]

\[ T_a \]

\[ T_p \]

\[ T_e \]

\[ T_0 \]

\[ t_p = 43.00\% \]

\[ t_e = 50.00\% \]

\[ t_a = 12.00\% \]

\[ t_c = 90.00\% \]

\[ T_0 = 100.00\% \]

**LF-parameters:**

\[ R_a = T_a/T_0 \]

\[ R_g = T_0/(2T_p) \]

\[ R_k = (T_e - T_p)/T_p \]

\[ OQ = T_e/T_0 = (1+R_k)/(2R_g) \]

**Basic shape parameter:**

\[ R_d = (T_0/T_0)(1/110) = \]

\[ = (U_0/E_e)(F_0/110) = \]

\[ \approx (0.5 + 1.2R_k)(R_k/(4R_g) + R_a)/0.11 \]

---

THE ROLE OF GLOTTAL SOURCE PARAMETERS FOR HIGH-QUALITY TRANSFORMATION OF PERCEPTUAL AGE

Xavier Favory, Nicolas Obin, Gilles Degottex, Axel Roebel

IRCAM - UMR STMS IRCAM-CNRS-UPMC
Paris, France
Voice source model

piecewise (exponential) function: discontinuity

discretization $\rightarrow$ aliasing

Severe degradation!
Before and after antialiasing

L-F model direct discretization

equalized antialiased L-F model

frequency (kHz)

time (s)

direct digitization

Antialiasing+digitization +equalization
Before and after antialiasing

L-F model direct discretization

equalized antialiased L-F model

direct digitization

Antialiasing+digitization +equalization
Before and after antialiasing

direct digitization

Antialiasing+digitization +equalization
Voice source model

Piecewise (exponential) function: discontinuity

discretization → aliasing

Severe degradation!
A. Aliasing

parameters [15] have been investigated for decades. Also, perceptual effects of these and methods for estimating these parameters have been studied [12], [13], [5], [14].

The source signal generated by the L-F model consists of complex exponentials and easily and efficiently represented by a cosine series defined below.

\[
E(t) = \sum_{n=-\infty}^{\infty} a_n e^{j2\pi f_0 t} + \sum_{n=1}^{\infty} \frac{b_n}{n} \sin(n\pi f_0 t),
\]

where \(a_n, b_n\) are coefficients of the cosine series.

B. Closed-form of band-limited source model

A set of closed-form representations for constituent piecewise polynomial model of the glottal source signal was proposed and derived three types of closed-form representations of its band-limited versions [10]. The formulation is the main contribution of this article. For speech synthesis.

This aliasing effects are severe when the sampling frequency is relatively low. Figure 2 shows an example. It shows the magnified view of the power spectrum of a L-F model signal, where the components other than \(f_0\) in the figure are caused by aliasing when sampled for discretization.

The time window function used in this analysis is designed using the discrete prolate spheroidal wave function [16] to assure low side lobe levels (lower than -60 dB). The best time-bandwidth product for finite support is calculated at \(\pi \leq \omega < 2\pi\). It is relatively low. Figure 2 shows an example. It shows the magnified view of the power spectrum of a L-F model signal, which consists of complex exponentials and easily and efficiently represented by a cosine series defined below.

\[
E(t) = \sum_{n=-\infty}^{\infty} a_n e^{j2\pi f_0 t} + \sum_{n=1}^{\infty} \frac{b_n}{n} \sin(n\pi f_0 t),
\]

where \(a_n, b_n\) are coefficients of the cosine series.

III. Anti-aliasing parameters

A piecewise polynomial model of the glottal source signal is derived for the L-F model and cosine series LPF functions [18]. The resulted representation only consists of spurious components are caused by aliasing. The time window function used in this analysis is designed using the discrete prolate spheroidal wave function [16] to assure low side lobe levels (lower than -60 dB). The best time-bandwidth product for finite support is calculated at \(\pi \leq \omega < 2\pi\). It is relatively low. Figure 2 shows an example. It shows the magnified view of the power spectrum of a L-F model signal, which consists of complex exponentials and easily and efficiently represented by a cosine series defined below.

\[
E(t) = \sum_{n=-\infty}^{\infty} a_n e^{j2\pi f_0 t} + \sum_{n=1}^{\infty} \frac{b_n}{n} \sin(n\pi f_0 t),
\]

where \(a_n, b_n\) are coefficients of the cosine series.
IV. ALTERNATIVE IMPLEMENTATION

Over-sampling and anti-aliasing filtering on the over-sampled time domain is a common practice. Figure 8 shows spectrograms using this practice. The over-sampled signal is using 48000 Hz sampling frequency, six times over-sampling. The upper spectrogram is calculated from the over-sampled L-F model output directly. The lower spectrogram is calculated from the over-sampled volume velocity signal based on L-F model. Appendix-B provides a set of equations to represent the L-F model-based volume velocity. The anti-aliasing LPF is designed on 48000 Hz sampling system and applied to the over-sampled signal. Then, the anti-aliased L-F model signal is down-sampled. For the volume velocity signal, differentiation is applied to the down-sampled signal.

Comparison of Fig. 7 and Fig. 8, illustrates that levels of spurious components are substantially reduced by over-sampling, especially for the volume velocity signal. Figure 9 provides support of this observation quantitatively. It shows the temporally averaged spectrograms. It also consists of the result of the direct L-F model discretization. The direct discretization introduces spurious level around -23 dB to the fundamental component. This spurious level is substantially reduced to about -50 dB using over-sampling and anti-aliasing filtering in the over-sampled domain. By using the volume velocity signal for over-sampling, the level is further reduced to about -70 dB. For the anti-aliased closed-form, the spurious level is around -120 dB. This is a substantial reduction.

In this figure, the spectrogram for the anti-aliased signal is calculated using the self convolution version of the Nuttall window mentioned before. By using self convolution technique recursively, the attenuation of side lobe levels are doubled each iteration. Note that self convolution of a cosine series window is also a cosine series. It means that the spurious levels of closed-form anti-aliasing signals can be arbitrarily suppressed using this self convolution technique.
Overview

- Interactive environment for speech science education
- Relation between vocal tract shape, transfer function, pole locations, LSF, F0, vibrato, voice source and speech sounds.
- A closed-form representation of the anti-aliased version of the well known voice source model (L-F model) is introduced
- Matlab-based implementation
Voice source panel
Conclusion

• Interactive environment for speech science education

• Relation between vocal tract shape, transfer function, pole locations, LSF, F0, vibrato, voice source and speech sounds.

• A closed-form representation of the anti-aliased version of the well known voice source model (L-F model) is introduced

• Matlab-based implementation

Please download and have fun!
Matlab realtime speech tools

by Hideki Kawahara

Matlab codes stored in this page are free to modify and use, since they are elemental and educational. All codes are provided "AS IS." However, I appreciate your comments and feedback. (Please refer to pp.5-11 of APSIPA Newsletter Issue9.)

In old Matlab versions, "audiorecorder" object does not allow data acquisition while it is running, and the tools in this page do not work. I appreciate your comments.

(Please mail me, if you have any suggestions. I appreciate.)

Matlab code

- Preliminary release of a vocal tract shape to voice synthesis GUI with an anti-aliased L-F model source. (Matlab source codes)
  (01/August/2015:revised)
  Introduction to this tool will be presented at a technical meeting of IEICEJ on 3rd August, 2015. (See the technical report below.)
  - Link to the technical report of IEICEJ (3/Aug./2015)
  - Presentation slides (PDF: 6MB)(3/Aug./2015)
  - Quick tour (movie: 60MB, 6 minutes)
  - Soprano voice simulation (movie: 16MB, 90 seconds)
  - Short movie fragments: (Short demo) (LF model designer demo)

- The following installers (01 August, 2015 version) download big files (approx. 750MB) from Mathworks
  - Stand alone application for Mac OSX (64bit) [installer]
  - Stand alone application for Windows (64bit) [installer]

These do not require Matlab
Equalizer

upper limit: 50 (dB)