Exploration of the other aspect of Vocoder revisited

Hideki Kawahara
Wakayama University, Japan
I am a tool builder hoping to make useful tools to promote understanding of human speech communication and to encourage collaborations between researchers and developers. I would appreciate your suggestions for me to produce further interesting tools.
Outline

• Background: Vocoder
  • Why are voiced sounds periodic?
• TANDEM: temporally static representation
• STRAIGHT: F0-adaptive spectral smoothing with compensation
• XSX: exhaustive periodicity detector
• Temporally variable multi-aspect morphing: a tool for explorations
• Challenge everywhere: aperiodicity
TANDEM-STRAIGHT: periodic pulse power spectrum
TANDEM-STRAIGHT: synthetic vowel /a/

log-power spectrum

Short time Fourier transform (STFT)

frequency (Hz)

0 500 1000 1500 2000 2500 3000

0 -10 -20 -30

time (s)

0.03 0.04 0.05

0 0.05

0 -0.05

2010年9月24日金曜日
Outline

• Background: Vocoder
  • Why are voiced sounds periodic?
• TANDEM: temporally static representation
• STRAIGHT: F0-adaptive spectral smoothing with compensation
• XSX: exhaustive periodicity detector
• Temporally variable multi-aspect morphing: a tool for explorations
• Challenge everywhere: aperiodicity
TANDEM: principle

- Static representation of power spectrum

**signal model**

\[ x(t) = e^{jk\omega_0 t} + ae^{j((k+1)\omega_0 + \beta)} \]

**power spectrum**

\[ P(\omega, t) = |W(\omega)|^2 + \alpha^2 |W(\omega - \omega_0)|^2 + 2W(\omega)W(\omega - \omega_0) \cos(\omega_0 t + \beta) \]

**windowing function**

**temporally varying term**

**TANDEM spectrum**

\[ P_T(\omega, t) = \frac{1}{2} \left[ P \left( \omega, t - \frac{T_0}{4} \right) + P \left( \omega, t + \frac{T_0}{4} \right) \right] \]
Implementation

● How to select windowing function

\[ \eta_t = \frac{\sqrt{\int \int_0^{T_0} |P(\omega, t) - \overline{P(\omega)}|^2 \, dt \, d\omega}}{\int \int_0^{T_0} P(\omega, t) \, dt \, d\omega} \]

- temporal variation
- normalized duration

\[ \sigma_t = \frac{1}{T_0} \sqrt{\int_{-T_w/2}^{T_w/2} t^2 w^2(t) \, dt} \]

- averaged spectrum

2010年9月24日金曜日
temporal variation

single window

normalized duration (re. T0)
temporal variation
TANDEM window
normalized duration (re. T0)
temporal variation

TANDEM window

20dB

single window

normalized duration (re. T0)

normalized effective window length (re. T0)
Implementation

• Which window function to use?

\[ \eta_{dBt} = \sqrt{\left\langle \frac{1}{2\pi T_0} \int_0^T \int_0^{T_0} \left| L(\omega, t) - \overline{L(\omega)} \right|^2 dt \ d\omega \right\rangle} \]

temporal variation of log-power spectrum

\[ \overline{L(\omega)} = \left\langle \frac{1}{T_0} \int_0^{T_0} L(\omega, t) \ dt \right\rangle, \quad L(\omega, t) = 10 \log_{10} P(\omega, t) \]

averaged log-power spectrum
temporal variation of log-power spectrum

normalized duration (re. T0)
temporal variation of log-power spectrum

normalized duration (re. T0)
temporal variation of log-power spectrum

single window

normalized duration (re. T0)

TANDEM window
Best window

- Blackman window with $2.5T_0$ base
- Perfect cancellation at harmonic frequencies and mid-points
- Shortest actual window length among windows with the same effective length
actual window length

normalized window length (re. $T_0$)

normalized effective window length (re. $T_0$)

effective window length

$hanning$
$blackman$
nuttall
$kaiser$

$2.5T_0$

original
TANDEM

2010年9月24日金曜日
Outline

• Background: Vocoder
  • Why are voiced sounds periodic?
• TANDEM: temporally static representation
• STRAIGHT: F0-adaptive spectral smoothing with compensation
• XSX: exhaustive periodicity detector
• Temporally variable multi-aspect morphing: a tool for explorations
• Challenge everywhere: aperiodicity
Eliminating frequency domain periodicity

- smoothing function having zeroes at spatial harmonic frequencies → rectangular window with width: $\omega_0$

\[
PS(\omega, t) = \exp \left( \frac{1}{\omega_0} \int_{-\frac{\omega_0}{2}}^{\frac{\omega_0}{2}} \log(P_T(\omega - \lambda))d\lambda \right)
\]

\[
PS(\omega) \approx P_{TST}(\omega)
\]
Consistent sampling

Is sampling theory relevant for envelope reconstruction?

Consistent sampling: recovery only at sampled points

Sampling theory: whole waveform recovery
Consistent sampling

- Is sampling theory relevant for envelope reconstruction?

Sampling theory: whole waveform recovery

Consistent sampling: recovery only at sampled points

simple filtering
Consistent sampling

- Digital filter design for compensation

\[ P_{ST}(\omega, t) = \sum_{k=-\infty}^{\infty} q_k P_S(\omega - k\omega_0, t) \]

\[ Q(z) = \frac{1}{R(z)} = \frac{1}{\sum_{k=-\infty}^{\infty} r_k z^{-k}} = \sum_{k=-\infty}^{\infty} q_k z^{-k} \]

\[ r_k = \int_{-\infty}^{\infty} h(\omega - k\omega_0) |W(-\omega)|^2 d\omega, \]

- Recovered spectrum
- Smoothed spectrum
- Frequency domain representation
- Smoothing function
Consistent sampling

- simplification

\[ P_{ST}(\omega, t) = \sum_{k=-1}^{1} q_k P_S(\omega - k\omega_0, t) \]
Implementation

- Cepstrum-based liftering

truncated and
adjusted coefficients

cepstrum representation of TANDEM spectrum

\[
P_{TST}(\omega) = \exp \left( \mathcal{F}^{-1} \left[ \left( \hat{q}_0 + 2\hat{q}_1 \cos \left( \frac{2\pi \tau}{T_0} \right) \right) g(\tau) C(\tau) \right] \right),
\]

where

\[
g(\tau) = \frac{\sin(\pi f_0 \tau)}{\pi f_0 \tau} = \mathcal{F}[h_2(\omega)],
\]
Test example

- TANDEM and STRAIGHT spectra
Test example

- TANDEM and STRAIGHT spectra

![Graph showing TANDEM and STRAIGHT spectra with a question mark indicating a potential contradiction.]
TANDEM-STRAIGHT: natural speech
Outline

• Background: Vocoder
  • Why are voiced sounds periodic?
• TANDEM: temporally static representation
• STRAIGHT: F0-adaptive spectral smoothing with compensation
• XSX: exhaustive periodicity detector
• Temporally variable multi-aspect morphing: a tool for explorations
• Challenge everywhere: aperiodicity
Periodicity detection by spectral division
What is special?
TANDEM

STRAIGHT

F0 adaptive processing

sp. division

shaping
F0 adaptive processing

Contradiction:
no F0 information

multiple hypothesis and integration
Multiple hypothesis and integration

signal

detector-1

detector-2

detector-3

detector-4

\[ \sum \]

integration

F0 salience

normalized lag in semitone (re. T0)

response

blackman 2.5 npo:3 std:0.0012663

2010年9月24日金曜日
Spectral division

- TANDEM spectrum:
  <- envelope and periodic structure

- F0-adaptive smoothed spectrum
  <- envelope

\[ PP(\omega) = \frac{PT(\omega)}{P_{TST}(\omega)} - 1 \]

spectrum only with periodic component

smoothed spectrum
Selecting base-band

\[ r_A(\tau) = \int_{-\infty}^{\infty} w_B(\omega) P_P(\omega) e^{j\omega \tau} d\omega, \]

where \( w_B(\omega) = \begin{cases} c_B \left(1 + \cos\left(\frac{\pi \omega}{N\omega_0}\right)\right) & |\omega| \leq N\omega_0 \\ 0 & |\omega| > N\omega_0 \end{cases} \)
Integration of individual detectors

\[ r(\tau; f_c) = w_L(\tau; f_c) r_A(\tau; f_c), \]

where \[ w_L(\tau; f_c) = \begin{cases} 
1 + \cos(\pi u(\tau)) & |u(\tau)| \leq 1 \\
0 & |u(\tau)| > 1 
\end{cases} \]

\[ u(\tau) = b_w \log_2(\tau f_c), \]

\[ r_I(\tau) = c_0 \sum_{f_c \in F_c} r(\tau; f_c), \]
Integration of individual detectors

channels: 2  bw: 2  compensators: [0.075, -0.2]

center frequency: 40 (Hz)
frame#: 90
Integration of individual detectors

\[ r(\tau; f_c) = w_L(\tau; f_c) r_A(\tau; f_c), \]

where

\[ w_L(\tau; f_c) = \begin{cases} 
1 + \cos(\pi u(\tau)) & |u(\tau)| \leq 1 \\
0 & |u(\tau)| > 1
\end{cases} \]

\[ u(\tau) = b_w \log_2(\tau f_c), \]

\[ r_I(\tau) = c_0 \sum_{f_c \in F_c} r(\tau; f_c), \]
Integration of individual detectors

channels: 2, bw: 2, compensators: [0.075, -0.2]

center frequency: 226 (Hz), frame#: 5001

shaped response
integrated response
Alternating amplitude

Salience of periodicity with candidates

frequency (Hz)

10^2

10^3

time (ms)
Displacement of pulse timing
Salience of periodicity with candidates

2010年9月24日金曜日
Salience of periodicity with candidates

frequency (Hz)  \(10^2\)  \(1660\) time (ms)

2010年9月24日金曜日
Salience of periodicity with candidates

F0

F0/5
Very fast F0 extractor
Outline

- Background: Vocoder
  - Why are voiced sounds periodic?
- TANDEM: temporally static representation
- STRAIGHT: F0-adaptive spectral smoothing with compensation
- XSX: exhaustive periodicity detector
- Temporally variable multi-aspect morphing: a tool for explorations
- Challenge everywhere: aperiodicity
Morphing voice identity and style

- re-formulation of morphing
Breakdown-free extrapolation

interpolation/extrapolation of logarithmic derivative
of coordinate transformation function

\[ T_{Am}(x_A) = \int_0^{x_A} \exp \left( \log \left( \frac{dT_{Am}(\lambda)}{d\lambda} \right) \right) d\lambda \]

\[ = \int_0^{x_A} \exp \left( (1 - r_{AB}(\lambda)) \log \left( \frac{dT_{AA}(\lambda)}{d\lambda} \right) \right) d\lambda \]

\[ + r_{AB}(\lambda) \log \left( \frac{dT_{AB}(\lambda)}{d\lambda} \right) d\lambda \]

\[ = \int_0^{x_A} \left( \frac{dT_{AB}(\lambda)}{d\lambda} \right)^{r_{AB}(\lambda)} d\lambda , \]
Morphing of temporal axis

- Incremental realtime update for interactive applications

\[ t_s = \int_0^{t_s} d\lambda, \]

\[ T_{SA}(t_s) = \int_0^{t_s} \left( \frac{dT_{AB}(T_{SA}(\lambda))}{d\lambda} \right)^{-r_{AB}^{(t)}(\lambda)} d\lambda, \]

\[ T_{SB}(t_s) = \int_0^{t_s} \left( \frac{dT_{BA}(T_{SB}(\lambda))}{d\lambda} \right)^{(r_{AB}^{(t)}(\lambda)-1)} d\lambda, \]
Morphing of temporal axis

- Formulation for post production

\[ \int_0^{t_s} \omega_s(\lambda) d\lambda \quad \text{excitation pulse} \]
Morphing of temporal axis

- Formulation for post processing
GUI tool (demo)
Outline

• Background: Vocoder
  • Why are voiced sounds periodic?
• TANDEM: temporally static representation
• STRAIGHT: F0-adaptive spectral smoothing with compensation
• XSX: exhaustive periodicity detector
• Temporally variable multi-aspect morphing: a tool for explorations
• Challenge everywhere: aperiodicity
I am a tool builder hoping to make useful tools to promote understanding of human speech communication and to encourage collaborations between researchers and developers. I would appreciate your suggestions for me to produce further interesting tools. Thank you!