Beyond band-limited sampling of speech spectral envelope imposed by the harmonic structure of voiced sounds

Hideki Kawahara\textsuperscript{1}, Masanori Morise\textsuperscript{2}, Tomoki Toda\textsuperscript{3}, Ryuichi Nisimura\textsuperscript{1} and Toshio Irino\textsuperscript{1}

\textsuperscript{1}Wakayama University, Japan
\textsuperscript{2}University of Yamanashi, Japan
\textsuperscript{3}Nara Advanced Institute of Science and Technology, Japan
Conclusion

Problem to be solved
- Recovery of spectral envelope details
- Vocal tract transfer function (shape) is not band-limited
- Periodic excitation of voiced sounds samples underlying spectral shape in the frequency domain

Solution
- Smoothing function with localized support
- Hybrid approach using vocal tract transfer function model as additional source of (missing) information

Further issues
- Estimation bias by non-Gaussian excitation
- Subjective evaluation
Outline

- Problem to be solved
  - Recovery of spectral envelope details
  - Vocal tract transfer function (shape) is not band-limited
  - Periodic excitation of voiced sounds samples underlying spectral shape in the frequency domain

- Solution
  - Smoothing function with localized support
  - Hybrid approach using vocal tract transfer function model as additional source of (missing) information

- Further issues
  - Estimation bias by non-Gaussian excitation
  - Subjective evaluation
Vocal tract transfer function
Frequency sampling by periodic excitation
Transfer function is not band-limited

\[ H(f) = \prod_{k=1}^{N} \left| \frac{1}{1 - z_k z^{-1}} \right| \]

\[ z = e^{j2\pi f}, \quad j = \sqrt{-1} \]

Gain (dB)

Frequency (Hz)
Morlet wavelet

\[ \psi(f) = \left( e^{j k_0 f} - e^{-\frac{k_0^2}{2}} \right) e^{-\frac{f^2}{2}} \]
Dilation: scaling factor $\rightarrow$ F0

Translation: frequency

\[
W_{\psi,H}(a,b) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{\infty} \psi^* \left( \frac{f-b}{a} \right) H(f) df
\]

\[
C(a,b) = M \left( 20 \log_{10} |W_{\psi,H}(a,b)| \right)
\]

Pseudo color mapping
dilation: scaling factor $\rightarrow F_0$

translation: frequency

\[ W_{\psi, H}(a, b) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{\infty} \psi^*(\frac{f-b}{a}) H(f) df \]

\[ C(a, b) = M(20 \log_{10}|W_{\psi, H}(a, b)|) \]

pseudo color mapping
Transfer function is not band-limited

\[ H(f) = \prod_{k=1}^{N} \left| \frac{1}{1 - z_k z^{-1}} \right| \]

\[ z = e^{j2\pi f}, \quad j = \sqrt{-1} \]
Wavelet transform of TF
Wavelet transform of TF
Outline

- Problem to be solved
  - Recovery of spectral envelope details
  - Vocal tract transfer function (shape) is not band-limited
  - Periodic excitation of voiced sounds samples underlying spectral shape in the frequency domain

- Solution
  - Smoothing function with localized support
  - Hybrid approach using vocal tract transfer function model as additional source of (missing) information

- Further issues
  - Estimation bias by non-Gaussian excitation
  - Subjective evaluation
waveform

TANDEM process

TANDEM spectrum

STRAIGHT process

STRAIGHT spectrum

spectral correction

extended STRAIGHT spectrum

auto correlation

inverse Fourier transform

LPC estimation

LPC spectrum

TANDEM and STRAIGHT smoothing

smeared LPC spectrum

spectral division

correction spectrum
Power spectrum
Wavelet transform of power spectrum

harmonic structure: 145Hz
Outline

- Problem to be solved
  - Recovery of spectral envelope details
  - Vocal tract transfer function (shape) is not band-limited
  - Periodic excitation of voiced sounds samples underlying spectral shape in the frequency domain

- Solution
  - Smoothing function with localized support
  - Hybrid approach using vocal tract transfer function model as additional source of (missing) information

- Further issues
  - Estimation bias by non-Gaussian excitation
  - Subjective evaluation
The diagram illustrates the process and steps involved in transforming a waveform into an extended STRAIGHT spectrum. The steps include:

1. **Waveform**
2. **TANDEM process**
3. **TANDEM spectrum**
4. **STRAIGHT process**
5. **STRAIGHT spectrum**
6. **Spectral correction**
7. **Spectral division**
8. **Correction spectrum**
9. **Auto correlation**
10. **Inverse Fourier transform**
11. **LPC estimation**
12. **LPC spectrum**
13. **TANDEM and STRAIGHT smoothing**
14. **Smeared LPC spectrum**
Smoothing function with localized support
Wavelet transform of smoothed spectrum
Wavelet transform of smoothed spectrum
waveform

TANDEM process

TANDEM spectrum

STRAIGHT process

STRAIGHT spectrum

spectral correction

extended STRAIGHT spectrum

inverse Fourier transform

auto correlation

LPC estimation

LPC spectrum

TANDEM and STRAIGHT smoothing

smeared LPC spectrum

spectral division

correction spectrum
LPC-spectral envelope using STRAIGHT
Wavelet transform of LPC spectrum
waveform

TANDEM process

TANDEM spectrum

STRAIGHT process

STRAIGHT spectrum

spectral correction

extended STRAIGHT spectrum

auto correlation

inverse Fourier transform

LPC estimation

LPC spectrum

TANDEM and STRAIGHT smoothing

smeared LPC spectrum

spectral division

correction spectrum
Enhanced STRAIGHT spectrum
Wavelet transform of smoothed spectrum

![Wavelet Transform Diagram](image)
Wavelet transform of enhanced STRAIGHT spectrum
Demonstration

- Original
- TANDEM-STRAIGHT with Rectangular smoother
- TANDEM-STRAIGHT with Triangular smoother
- Enhanced STRAIGHT with Rectangular smoother
- Enhanced STRAIGHT with Triangular smoother
- TANDEM-STRAIGHT with cepstrum lifter-based correction
Outline

- **Problem to be solved**
  - Recovery of spectral envelope details
  - Vocal tract transfer function (shape) is not band-limited
  - Periodic excitation of voiced sounds samples underlying spectral shape in the frequency domain

- **Solution**
  - Smoothing function with localized support
  - Hybrid approach using vocal tract transfer function model as additional source of (missing) information

- **Further issues**
  - Estimation bias by non-Gaussian excitation
  - Subjective evaluation
Conclusion

- Problem to be solved
  - Recovery of spectral envelope details
  - Vocal tract transfer function (shape) is not band-limited
  - Periodic excitation of voiced sounds samples underlying spectral shape in the frequency domain

- Solution
  - Smoothing function with localized support
  - Hybrid approach using vocal tract transfer function model as additional source of (missing) information

- Further issues
  - Estimation bias by non-Gaussian excitation
  - Subjective evaluation

Thank you!
Subjective evaluation: preliminary

- Japanese vowel sequence
- Speaker: male adult, female adult
- Procedure: paired comparison 2AFC
- Stimuli
  - Original, T-S with R, T-S with T, E-T-S with R, E-T-S with T, C-T-S
  - F0 modification, 0.7*F0, 1.0*F0, 1.3*F0
- Listener: 3 male, 2 female
- Analysis: GLM, binomial
- Result:
  - LPC-enhancement, highly significant, Cepstrum, highly significant