Inharmonic Speech: A Tool for the Study of Speech Perception and Separation

1. What is inharmonic speech?
2. Why make inharmonic speech?
3. How to make inharmonic speech
4. Psychoacoustic experiments
The Structure of Speech

- Classic source/filter model
Harmonic Speech

- Periodic source pulses as a Fourier series:

\[
\sum_{n=-\infty}^{\infty} \delta(t - n\tau) = \frac{1}{\tau} \left( 1 + \sum_{k=1}^{\infty} 2 \cos \frac{2\pi k}{\tau} t \right)
\]
Inharmonic Speech

\[ f_n = n f_0 \]
\[ f_{n+1} - f_n = f_0 \]

\[ f_n = n f_0 + a f_0 \]
\[ f_{n+1} - f_n = f_0 \]

\[ f_n = n f_0 + b(n^2 - n) f_0 \]
\[ f_{n+1} - f_n = (1 + 2bn) f_0 \]

\[ f_n = n f_0 + cr_n f_0 \quad r_n \in [-1 \ldots 1] \]
\[ f_{n+1} - f_n = (1 + c\Delta r_n) f_0 \]

\[ s(t) = \sum_{n=1}^{N} \cos 2\pi f_n t \]
Why Inharmonic Speech?

- **Harmonicity** is believed important
  - .. to the **fusion** of sounds in auditory organization
  - .. for pitch perception (prosody, speaker identity)

- **Voiced speech has...**
  - multiple (resolved) harmonics = “**sparse**” spectrum
  - .. with similar **modulation** properties
  - .. in a **harmonic pattern**

- **How important is the “**harmonic pattern**”?**
  - See how well people (& machines)
    can organize and separate inharmonic speech
  - .. which is otherwise “natural”
  - maybe it’s enough to have a “**sparse**” spectrum?
Harmonicity for Separation

- Filtering of harmonics
  - after f0 is found
  
  Denbigh & Zhao 1992
  Avery Wang 1995

- Labeling of regions
  - by shared f0 candidate
  
  Brown 1992
  Hu & Wang 2004

- Input mixture
- Front end
- Signal features (maps)
- Object formation
- Discrete objects
- Grouping rules
- Source groups

Denbigh & Zhao 1992
Avery Wang 1995
Brown 1992
Hu & Wang 2004
Synthesizing Inharmonic Speech

- Based on STRAIGHT
  - decompose speech into:
    - $f_0$ (pitch track)
    - periodic envelope (voiced speech)
    - noise envelope (unvoiced speech component)
• STRAIGHT periodic source resynthesis
  ○ ... as individual pitch pulses
  ○ ... or as a set of Fourier components
    - which can be made inharmonic
Inharmonic Speech Properties

- Periodicity index calculated by Praat
  - histogram over 76 TIMIT utterances

- Pitch tracks calculated by Praat for an example utterance
Psychological Experiment

• Idea: See impact of removing harmonicity on ability to understand mixed words

• Listeners presented with one or two simultaneous words or utterances
  ○ measure accuracy at identifying all words
  ○ synthesized as harmonic, inharmonic, or whisper
Results

- **Harmonic** tokens a little easier to understand
  - but **inharmonic** tokens much better than whispered
  - different types of inharmonicity seem equivalent
  - **Spectral sparsity** is a big contributor to separation?
Conclusions

- **Harmonicity** of voice is thought to be important for auditory scene analysis
  - but hard to separate harmonicity and sparsity

- Modified STRAIGHT framework produces high-quality *inharmonic* tokens
  - excitation synthesized as sinusoids with arbitrary frequency tracks

- Preliminary experiments show that inharmonic tokens can still be separated
  - quantify contribution of harmonicity vs. sparsity