Converging Hearing and Speech Enhancement Technologies

Fan-Gang Zeng
University of California, Irvine
fzeng@uci.edu;
http://www.ucihs.uci.edu/hesp
Contents

• Chapter 1: The amazing speech chain
• Chapter 2: Speech recognition
• Chapter 3: Enhancement technologies
• Chapter 4: Industrial and research trends
The Speech Chain: From Production to Perception
The *broken* speech chain:
Opportunities for enhancement
Production:
Source : Filter = Fine structure : Envelope

The amazing speech

- Intensity: 128 dB SPL at a distance of 8ft. 2in
- Frequency range: 50 to 15,000 Hz
- Vocal pitch: 5 octaves!
Perception:
Peripheral and central processing

"c a t"

Cortex
Brainstem
Auditory nerve
Inner ear
Middle ear
Outer ear
Efferents
The amazing hearing

- Loudness: 0.5 nm sensitivity and a 120 dB dynamic range
- Pitch: 0.1% resolution for musicians
- Timing: 10 micro-sec resolution for localization

You can hear from whisper to 110-dB PA in a moment like this…
Cochlear implant status

- 200,000 users worldwide
- Performance = 80% in quiet
- Talk on the phone
- $1B revenue, $25B cap
- 3 FDA approved devices
- 10 start ups

- Power hungry
- Expensive
- Imperfect hearing: CI music vs. Original
Summary

- Chapter 1: Our work is interesting
Broadband spectrogram of the phrase "the top of the hill"
Temporal envelope cues

Speech pressure waveforms of six phrases uttered by a male adult speaker

Waveforms obtained by fullwave rectification and low-pass filtering at 20 Hz

- her chop
- her ship
- her mate
- her work
- her bull
- her pool

100 ms

- Rosen (1991) Royal Phil Soc Trans
Speech recognition with primarily temporal cues

Shannon, Zeng, Kamath et al. (1995) Science
Little math

- Flanagan (1980) “Parametric coding of speech spectra”

\[ s(t) \approx \sum_{k=1}^{N} A_k(t) \cos \left[ 2\pi f_{ck} t + 2\pi \int_{0}^{t} \dot{\phi}_k(\tau)d\tau + \theta_k \right] \]

- Discard absolute phase:

\[ s(t) \approx \sum_{k=1}^{N} A_k(t) \cos \left[ 2\pi f_{ck} t + 2\pi \int_{0}^{t} \dot{\phi}_k(\tau)d\tau \right] \]

- Discard relative phase (i.e., frequency modulation):

\[ s(t) \approx \sum_{k=1}^{N} A_k(t) \cos[2\pi f_{ck} t] \]
What is fine structure?

- Hilbert’s definition: Temporal envelope and Fine structure

Original

AM

Fine Structure

FM
Implementation

- Combo of Dudley’s vocoder and Flanagan’s phase vocoder

Zeng, Nie, Stickney et al. PNAS (2005)
Spectra: What does FM encode?

Zeng, Nie, Stickney et al.  
PNAS (2005)
Sentence, speaker, and tone recognition

Zeng, Nie, Stickney et al. PNAS (2005)

Combo:  
Target:  
Masker:  

Zeng, Nie, Stickney et al. PNAS (2005)
Role of common FM: Binding and segregation
A Unified Model

Production

Source

Filter

Acoustics

Fine structure

Envelope

Perception

Object

Intelligibility
Summary

• Chapter 1: Our work is interesting
• Chapter 2: Speech cues are redundant and complementary
Clear speech perception

Liu, Del Rio, Bradlow and Zeng, JASA (2005)
Speech recognition with hearing aid and cochlear implant

Kong, Stickney, and Zeng, JASA (2005)
Summary

• Chapter 1: Our work is interesting
• Chapter 2: Speech cues are redundant and complementary
• Chapter 3: Increasing functional signal-to-noise ratio is the key
Converging technologies: Hearing aid or Bluetooth headset

Turning iPhone into a hearing aid
Micro- and Nano-Technology

Audio Processor

Internal Receiver

Conductor Link

Floating Mass Transducer

UCI 2009
Inner ear powered radio
Applications for speech and hearing enhancement:

- Stuttering
- Aphasia
- Dyslexia
- Learning disability
- Multiple sclerosis
- Alzheimer’s disease
- Autism
  ...

![Healthy brain](image1)

![Brain with damage](image2)

![Autism](image3)
Summary

• Chapter 1: Our work is interesting
• Chapter 2: Speech cues are redundant and complementary
• Chapter 3: Increasing functional signal-to-noise ratio is the key
• Chapter 4: Converging technologies to solve multiple problems