2014 Speech Signal Processing Workshop
National Taipei University, Taipei

Speech perception in adverse listening conditions

(Felix) Fei Chen
Division of Speech & Hearing Sciences, The University of Hong Kong

Aug. 01, 2014

Outline

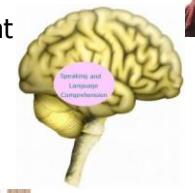
1. Background: Communication disorder
2. Experiments
   1. Tone contour contribution
   2. Vowel contribution
   3. Low frequency contribution
   4. Spectral contribution
3. Summary

Communication disorder

Communication disorders (CD) commonly refer to problems in speech, language and hearing (comprehension and/or expression) that significantly interfere with an individual's achievement and/or quality of life (Gleason, 2001).

Challenges to communication

- Hearing loss
- Voice problems
- Language impairment
- Autism
- Aphasia
- Stuttering
- …



Parkinson's disease

www.catalogs.com www.therapychd.com www.thelibertybeacon.com

Impact of CD study

- Improving the quality of life, and happiness of family and society

AMERICAN SPEECH-LANGUAGE-HEARING ASSOCIATION

Get Help..Don't Wait!

Effective Communication is a Human Right, Accessible and Achievable for ALL!

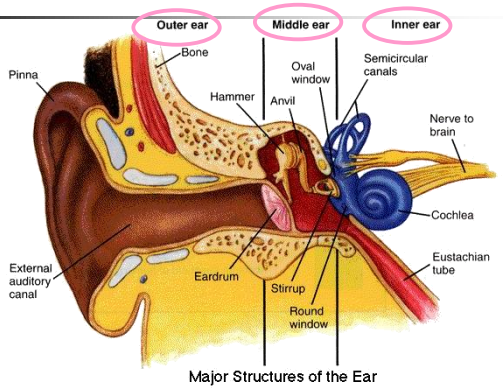
5

Hearing impairment: Facts

- The most frequent sensory deficit in human, affecting more than 250 million people in the world.
  - Reduced ability to communicate, delay in language acquisition, economic and educational disadvantage, social isolation, etc.
- Among the geriatric population, hearing impairment is the third most reported chronic condition (following arthritis and hypertension) (NIH report).
- Each year, about 5 million adolescents are diagnosed with noise-induced hearing loss around the word.
- …

6

Ear



Major Structures of the Ear

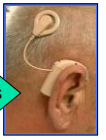
Image from Brown & Introductory Psychology Electronic Image Bank

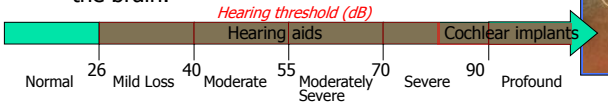
7

Assistive hearing devices

1. Hearing aids (HAs) and conductive hearing loss
   - When sound is not conducted efficiently through the outer and middle ears.
2. Cochlear implants (CIs) and sensorineural hearing loss
   - When there is damage to the inner ear (cochlea) or to the nerve pathways from the inner ear to the brain.

Images: drkirtane.com





8

Assistive hearing devices: Challenges

- Low-cost
- Speech perception in noise
- Music appreciation
- Tonal language (Mandarin and Cantonese) perception
- …

9

F0 contour in Mandarin

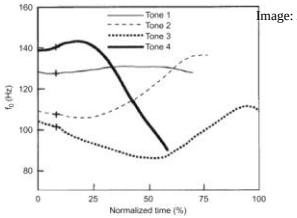


Image: Xu, 1997

"zhu"

Images: Yuen, et al. MAPPID-N

Language difference btwn English & Mandarin

1) Tone contour contribution: lexical meaning?
2) Syllable structure: mono-syllabic in Mandarin vs. multi-syllabic in English (e.g., 'ba' vs. 'manipulate');
3) Numbers of vowels/consonants: 35/21 of vowels/consonants in Mandarin vs. 20/32 in English (Chen et al., JASA, 2013; Fogerty et al., JASA, 2009); and
4) Proportion of vowel segments: 66% in Mandarin vs. 45% in English (Chen et al., JASA, 2013; Fogerty et al., JASA, 2009).

11

2. Experiments

To assess the language-specific contributions of F0 contour, vowels, and low frequencies to Mandarin sentence perception.

2.1 Tone contour contribution

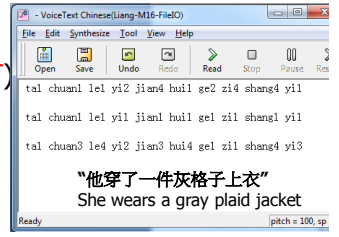
- Motivation
  - For HA/CI users, improvement of lexical tone identification does not lead to better Mandarin sentence perception (e.g., Milczynski et al., 2012).
- Hypothesis
  - The role of tone identification in sentence perception depends on listening conditions, e.g., in quiet or in noise.

13

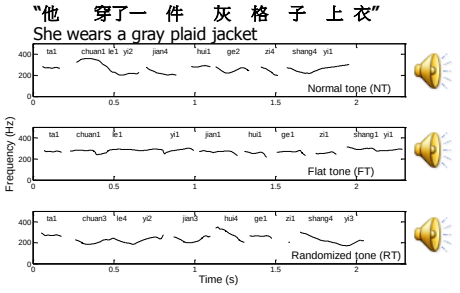
(cont.)

- Experiment design
  - Mandarin sentence synthesized by using high-quality text-to-speech (TTS) engine

1. Normal tone (NT)
2. Flat tone (FT)
3. Randomized tone (RT)

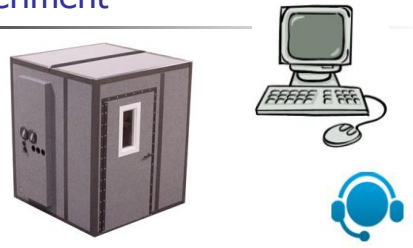


Examples



Chen et al., JSLHR, 2014

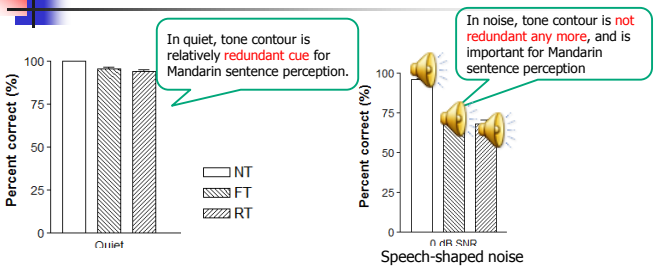
Experiment



- Subject: normal-hearing, native-Mandarin-speaking
- Procedure:
  - Training (10-20 mins), testing (break every 30 mins)
  - Score: How many words are correctly recognized

16

Results (n=11)

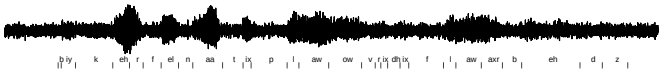


- Implication:
  - For assistive hearing devices, more efforts need to be focused on how to effectively deliver the tonal information in challenging listening conditions, e.g., in noise.

Chen et al., JSLHR, 2014

2.2 Vowel contribution

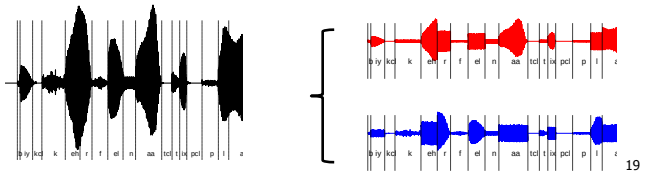
- Motivation
  - Noise has different influence to different phonetic segments (e.g., strong vowels and weak consonants).
  - Which segment contains more perceptual information? vowel or consonant segment?



18

Noise-replacement paradigm

- A remarkable (2:1) advantage of vowels (Vs) vs. consonants (Cs) for English sentence perception (Cole et al., 1996; Kewley-Port et al., 2007).
  - 87.4% (V-only sentences, with Cs replaced by white noise)
  - 46.6% (C-only sentences, with Vs replaced by white noise)



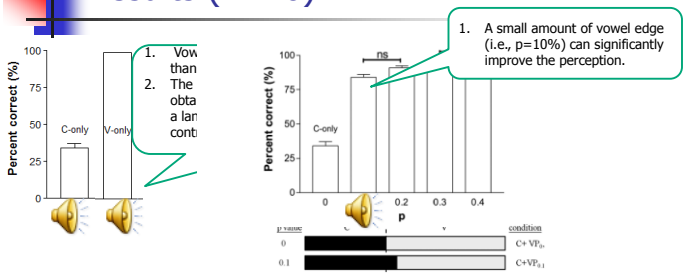
19

Examples

- English [She had your dark suit in greasy wash water all year]
  - V-sentence
  - C-sentence
- Mandarin [他剛才講話時吞吞吐吐]
  - V-sentence
  - C-sentence

20

Results (n=20)



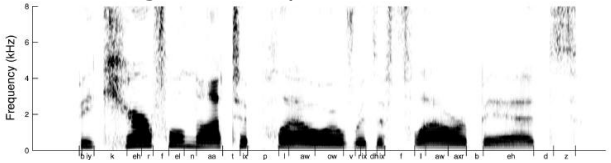
1. A small amount of vowel edge (i.e., p=10%) can significantly improve the perception.

- Implications:
  - Selectively highlight or train the vowel segments for speech perception.
  - Vowel-consonant transition carries important information for speech perception.

Chen et al., JASA, 2013

2.3 Low-frequency contribution

- Motivations
  - As vowels contain low-frequency components, how low-pass cutoff frequency affects the perception of low-frequency V-sentences?
  - Most hearing-impaired patients have residual hearing at low frequencies.

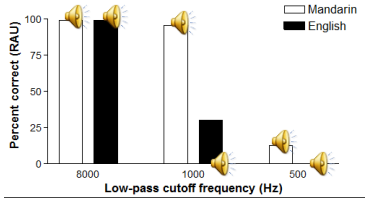


22

Results (n=8)

She had your dark suit in greasy wash water all year.

他剛才講話時吞吞吐吐



* Data collected from native-Mandarin and native-English subjects.

- Implications:
  - Mandarin vowels appear to be weighted towards low frequencies under 1000 Hz, which primarily contain F0 cue important for tone identification.
  - English vowels are much less effective than Mandarin vowels at low frequencies.

Chen et al., InterSpeech, 2014

2.4 Effect of spectral degradation

- Motivation
  - To which extent does acoustic cue (F0 contour, harmonics, formants, etc.) accounts for the perceptional advantage of vowels?

/eh/ in 'head'

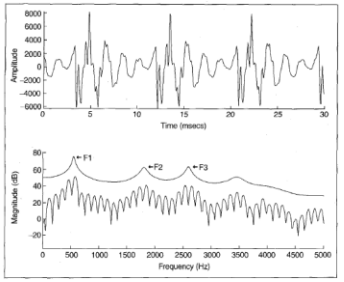


Image: Loizou, IEEE EMB, 1999.

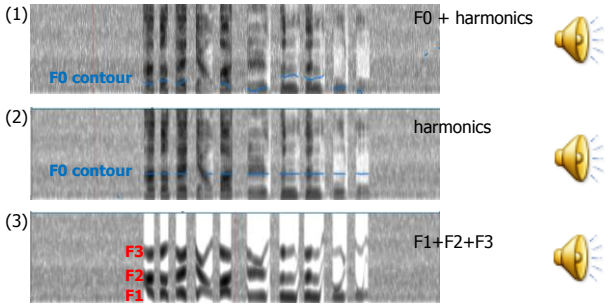
24

(cont.)

- Experiment design – 3 types of synthesized sentences
  1) V-sentence (as control);
  2) V-sentence with flattened F0 (no F0 contour but harmonics); and
  3) V-sentence containing only three formants (F1, F2, and F3), i.e. no F0 or harmonics.

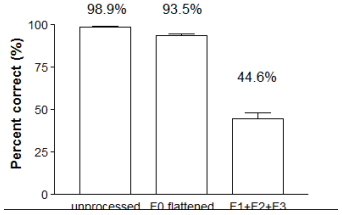
25

Examples



26

Results (n=20)



- F0 contour: a relatively smaller effect to the perception of Mandarin vowel sentences.
- Harmonics: carry more perceptual information.
- The first three formants: account much for speech perception.

Chen et al., InterSpeech, 2014

Summary

1) Language differences between English & Mandarin were revealed in contributions of vowels and low frequencies for sentence perception.
2) The contribution of F0 contour to Mandarin sentence perception could be compensated by other cues (e.g., top-down process), and is important in challenging listening conditions (e.g., in noise).
3) The vowel advantage is largely attributed to its harmonic structure, but not F0 contour.

28

Acknowledgements:
Faculty Research Fund, HKU; GRF Fund, Hong Kong RGC