Divergent patterns of voicing perception in various challenging listening conditions

Matthew B. Winn **William J. Idsardi** **Monita Chatterjee**

University of Maryland College Park

UMCP NACS PROGRAM NEUROSCIENCE AND COGNITIVE SCIENCE

CENTER FOR COMPARATIVE AND EVOLUTIONARY BIOLOGY OF HEARING

CEBH

Introduction

A phonetic contrast can be perceived using various acoustic cues.

Some acoustic cues are compromised by hearing impairment / limited bandwidth, background noise, or the use of a cochlear implant.

When faced with these challenging listening conditions, people can still hearing the voicing contrast very well. How do they do it?

Perhaps they can change the way they use different acoustic cues in the signal.

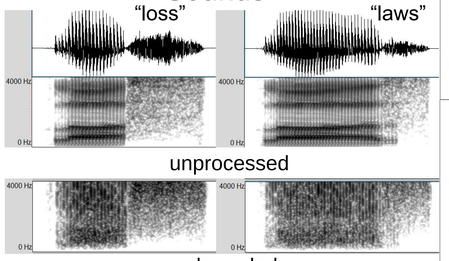
When we assess accuracy in speech perception, is it enough to keep track of words, phonemes and features correct?

Effects of Spectral Degradation

Problem: Poor spectral resolution obscures spectral cues such as formant transitions.
Do listeners in these conditions increase their use of durational cues when hearing voicing contrasts?

Methods

Sounds



Parameters

F1 offset
dropped to 415 Hz or remained at 615 Hz
Vowel duration
varied between 175 and 325 ms *(7 steps)*
Consonant duration
varied between 100 and 250 ms *(3 steps)*
Voicing duration
varied between 0 and 30 ms *(3 steps)*
Fully-crossed stimulus set

Task

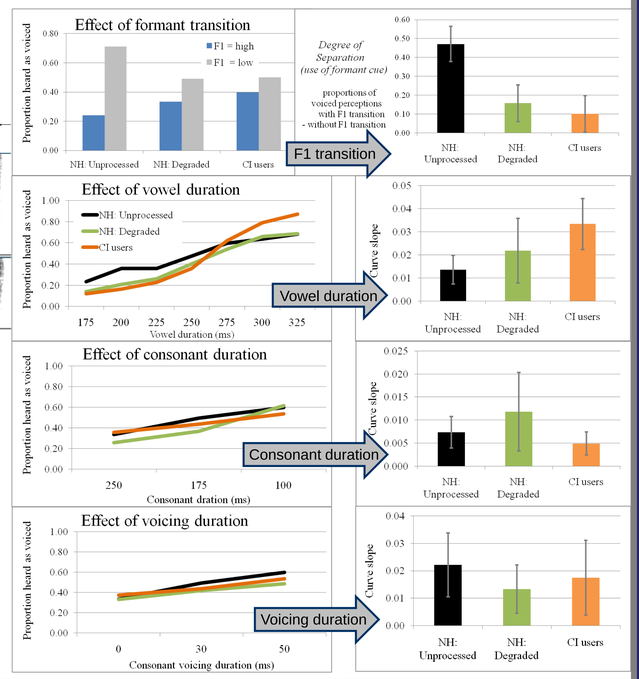
2- alternative forced-choice identification
"Loss" or "Laws"

Unprocessed (natural) or degraded with an 8-channel noise-band vocoder (Shannon et al., 1995)
Cochlear implant users only heard the unprocessed speech.

Analysis

Generalized linear mixed-effects model (binary link)

Psychometric function curve coefficient (slope) derived using standard sigmoidal function



GLMM results

F1 transition was used less when spectral resolution was degraded ($p < 0.001$)
F1 transition was used less by cochlear implant users ($p < 0.001$)

Vowel duration was *not* used significantly more when spectral resolution was degraded ($p = 0.127$)
Vowel duration was used more by cochlear implant users ($p < 0.001$)

Consonant voicing was used less by cochlear implant users ($p < 0.001$)

Vowel-consonant duration **ratio** was used more by cochlear implant users ($p < 0.05$)

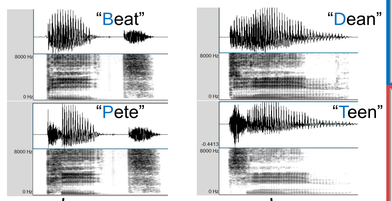
Consonant duration was not affected by spectral degradation or the use of a CI.

Effects of Bandwidth and Background Noise

Problem: these factors can obscure the voice-onset-time cue for voicing in stop sounds
Do listeners in these conditions increase their use of voice pitch (F0) cues when hearing voicing contrasts?

Methods

Sounds



Vowel onset used as reference for loudness of noise.
Noise: Speech shaped, steady-state.
Low-pass filtering: Hann filter using Praat.

Participants: 12 normal-hearing listeners

Parameters

Voice-onset-time: varied in 10ms steps
between -10 to +50 (b/p) and between 0 to 70 (d/t)
F0: onsets of 94, 100, 106, 112, 119, 126, 134, 142
Fully-crossed stimulus set

Task

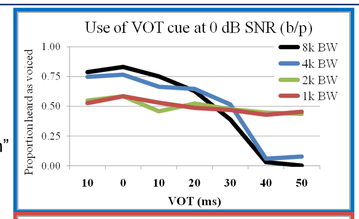
2- alternative forced-choice identification
"Beat" / "Pete" or "Teen" / "Dean"

In varying conditions of LPF and background noise

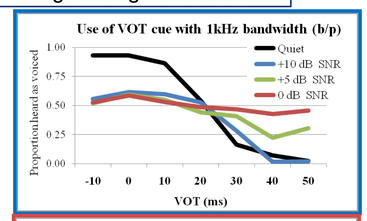
Analysis

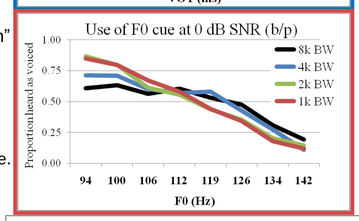
Generalized linear mixed-effects model (binary link)

Psychometric function curve coefficient (slope) derived using standard sigmoidal function

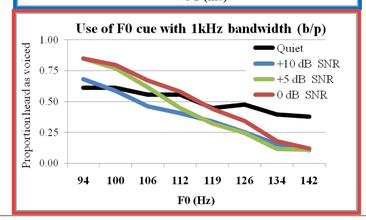


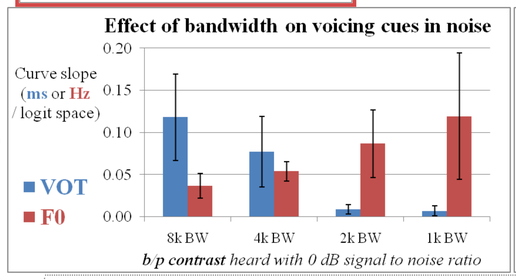
Decreasing effect of VOT as SNR or bandwidth becomes less favorable

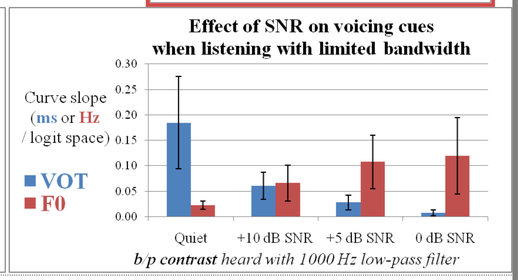




Increasing effect of F0 as SNR or bandwidth becomes less favorable







GLMM results

The use of **VOT** was affected by **SNR** for both contrasts ($p < 0.001$)
The use of **F0** was affected by **SNR** for both contrasts ($p < 0.001$);
SNR affected the b/p contrast more heavily than the d/t contrast.
The use of **VOT** was affected by **bandwidth** for d/t ($p < 0.001$), *but not for* b/p
The use of **F0** was affected by **bandwidth** for both contrasts ($p < 0.001$);
the d/t contrast was more heavily affected than the b/p contrast.
The use of **F0** was affected by **an interaction between SNR and bandwidth** for d/t ($p < 0.001$) *but not for* b/p ($p = 0.164$).
The use of **VOT** was affected by **an interaction between SNR and bandwidth** for both b/p and d/t.

Conclusions

1) With degraded spectral resolution (CI listeners or NH listeners in simulations) the formant transition cue played a smaller role, and temporal cues (vowel duration, consonant duration, vowel-consonant duration ratio) played a larger role in perception of final consonant voicing.

2) With background noise, voice-onset-time plays a smaller role, and F0 plays a larger role, especially for the b/p contrast.

3) With bandwidth reduction, voice-onset-time plays a smaller role, and F0 plays a larger role, especially for the d/t contrast.

4) Listeners in optimal conditions, with hearing impairment or with a cochlear implant can thus achieve the same performance on a speech recognition task (word recognition, phoneme recognition, confusion matrix/information transfer analysis), but through different perceptual / decision processes.

Many thanks go out to Qian-Jie Fu for his software, and to our participants

1. Shannon et al. (1995). Speech recognition with primarily temporal cues. Science, 270, 303–304.
2. Hillenbrand et al. (1984). "Perception of the voiced-voiceless contrast in syllable-final stops," JASA, **76,** 18-26.
3. Abramson and Lisker (1985). "Relative power of cues: F0 shift versus voice timing," In Phonetic Linguistics, (V. Fromkin)
4. Phatak and Allen (2007), "Consonant and vowel confusions in speech-weighted noise," JASA, **121**, 2312–2316.
5. Friesen et al. (2001). "Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants," JASA, **110,** 1150-1163.
6. Morrison (2005), "An appropriate metric for cue-weighting in L2 speech perception," SSLA, **27**, 597-606.

This research is supported by National Institutes of Health R01 DC 004786 (Monita Chatterjee) and T32 DC000046-17 (Arthur N. Popper).