The Importance of High-Frequency Information for Understanding "Glimpsed" Speech

Virginia Best, Elin Roverud, Lucas Baltzell, Jan Rennies & Gerald Kidd, Jr.
Department of Speech, Language & Hearing Sciences

Boston University College of Health
& Rehabilitation Sciences: Sargent College
Department of Speech, Language & Hearing Sciences

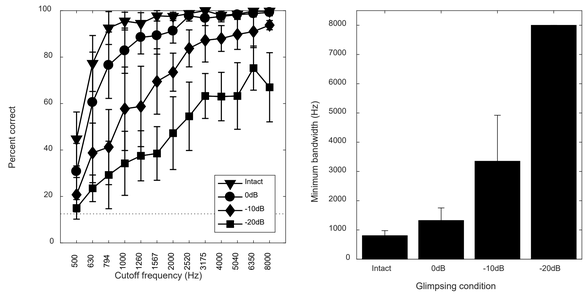
BACKGROUND

- When speech is interrupted by other talkers, listeners must not only segregate the voices but also recreate the target message from the available time-frequency "glimpses".
- Here we tested the hypothesis that high-frequency audibility is more important for sparse representations of speech than for intact speech.
- This question may be relevant for understanding the impact of high-frequency hearing loss on everyday speech communication.

METHODS

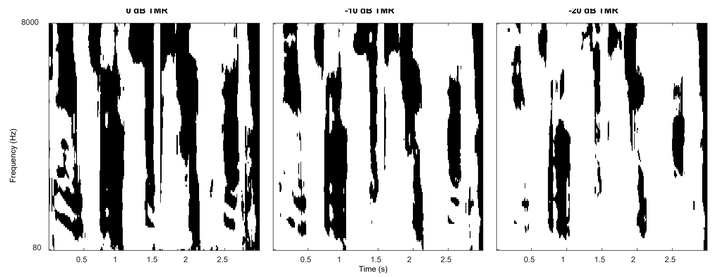
- Subjects were 8 healthy young adults with normal hearing (mean age 26 years).
- Stimuli were based on a mixture of sentences spoken by three different female talkers. For example:
  Target: Sue bought two red toys
  Masker1: Bob found six old socks
  Masker2: Pat lost five new pens
- The target sentence was presented in its **intact** form or was progressively **glimpsed** according to the two-talker masker presented at various levels (target-to-masker ratios of 0, -10, -20 dB).
- Intelligibility was measured for a range of low-pass conditions (cutoff frequencies from 500-8000 Hz).
- Stimuli were presented diotically over headphones and listeners responded by selecting 5 words from a grid of 40 possible words.

RESULTS



- Performance as a function of cutoff frequency showed clear differences across glimpsing conditions.
- The "minimum bandwidth" (Silberer et al 2015) for optimal performance rose from < 1 kHz (intact) to 8 kHz (most sparse).

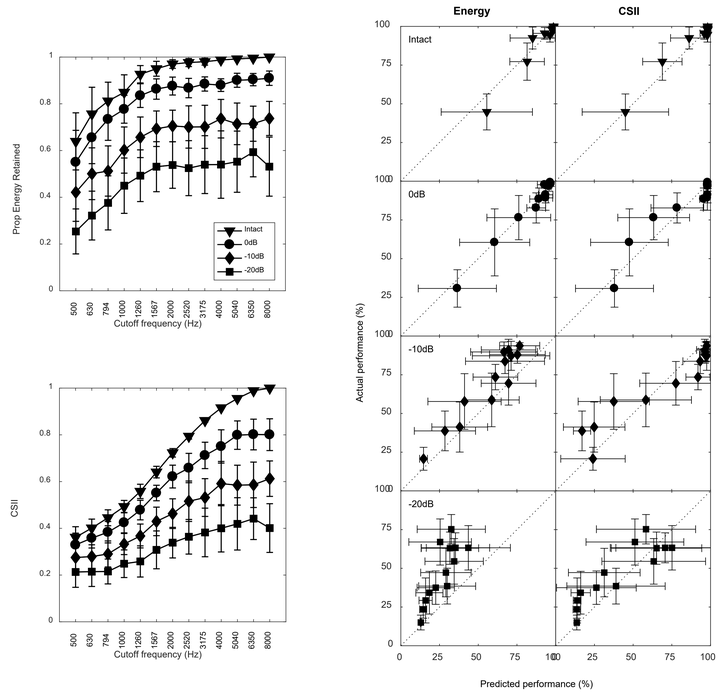
GLIMPSING MODEL



- Signals were analyzed using 128 frequency channels logarithmically spaced between 80 Hz and 8 kHz, and 20-ms time windows with 50% overlap.
- A binary mask was generated by assigning a value of 1 to time-frequency "tiles" in which the target energy exceeded the total masker energy, and a value of 0 to the remaining tiles.
- The mask was then applied to the clean target signal before resynthesis.
- Using this approach, the sparseness of the mask varies with TMR and in the resulting speech there are fewer glimpses at poorer TMRs.

ACOUSTIC ANALYSIS

- Acoustic analysis explored whether the combined effect of glimpsing and low-pass filtering could be understood in terms of available speech information.
- Two metrics applied to 50 random stimuli:
  1. Retained target energy (ratio of the RMS levels of the glimpsed/filtered stimulus and the intact/unfiltered stimulus).
  2. Coherence-based speech intelligibility index (CSII; Kates & Arehart 2005).
- While both metrics were reduced by glimpsing and low-pass filtering, the CSII was better able to capture the performance data.



CONCLUSION

- A broad bandwidth of speech information becomes increasingly important when speech is sparsely represented.
- A closer consideration of the audibility of high-frequency information may be needed to fully understand the difficulties experienced by many listeners in "cocktail party" situations.

REFERENCES

Kates JM & Arehart KH (2005). Coherence and the speech intelligibility index. JASA 117:2224–2237.

Silberer AB, Bentler R & Wu YH (2015). The importance of high-frequency audibility with and without visual cues on speech recognition for listeners with normal hearing. IJA 54:865-872.

Work supported by NIH-NIDCD award DC015760 to VB.