High Frequencies in Speech and Talker Localization and Segregation

Simon Carlile,
Department of Physiology,
University of Sydney

simonc@physiol.usyd.edu.au
Setting the scene

- The contribution of the various acoustic spatial cues to speech segregation and speech intelligibility?
- Particular focus on frequency bandwidth and the percept of externalisation.
- Using VAS and CRM to measure speech intelligibility with dichotic and virtual listening conditions.
• Significant high frequency content – e.g. “sludge”
• Not required for single channel speech intelligibility
  E.g. AI and SII models
• Frequencies > 8 kHz important in accurate localisation of bb noise. What about speech?
Localisation of speech stimuli
High frequencies & accurate speech localisation

Localisation Errors
  Lateral angle error
  Polar angle error

(a) No change in lateral angle error
(b) Significant increase in polar angle error (> 30°)

Jin et al 2002
Implications/questions.

- Differences in location an important cue in solving the cocktail party problem (Cherry)
- High frequencies important in accurate localisation of speech sources
- The Question: Does high frequency information in speech contribute to speech intelligibility with multiple concurrent talkers?
- The experiment: Measure speech reception threshold (SRT) with and without high frequency information
The Presentation Paradigm: Virtual Auditory Space

(a) Sound level over time

(b) Lateralised sounds
Free field listening

(c) Virtual Auditory Space
The intelligibility paradigm

Speech reception threshold measured using Coordinate Response Measure

Target  Target+maskers  Ready Baron go to blue one now
Talkers: 4 male 4 female

Adaptively vary level of target (Quest) to obtain the signal to masker threshold.

Listening model: The “Dinner table” problem

Auditory Neuroscience Laboratory, University of Sydney
Measuring Speech Intelligibility: the “dining table” model

TARGET

MASKER

MASKER

Level (dB)

INTELLIGIBILITY

TARGET

MASKER 1

+ MASKER 2

SRT

Release from masking

Scale

100%

0%
Diotic listening

Speech Reception Thresholds; Diotic presentation of sounds

-6 -5 -4 -3 -2 -1 0 1 2 3 4 5 6

Subjects

Level of Target with respect to Maskers (dB)

Mean -3.8dB

• No spatial separation of sound sources
• Internalisation of sound sources

Same result when co-located stimuli are externalised to directly ahead.

Auditory Neuroscience Laboratory, University of Sydney
Spatialised (VAS) listening

Speech Reception Thresholds;
VAS presentation of sounds
Maskers separated at (0,30) and (0,-30)

-25 -20 -15 -10 -5 0 1 2 3 4 5 6

Mean -19.4dB

- Sound sources spatially separated
- Sound sources perceived in external space
Binaural listening

<table>
<thead>
<tr>
<th>Condition</th>
<th>-20</th>
<th>-15</th>
<th>-10</th>
<th>-5</th>
<th>0</th>
</tr>
</thead>
</table>

- Sounds sources separated by differences in the binaural cues
- Sound sources lateralised within the head

Mean -13.4 dB
5 kHz low pass VAS listening

5 kHz Low-passed VAS presentation of sounds; maskers positioned at (0,30) and (0,-30)

- Subjects
- Level of Target with respect to Target (dB)

Mean -15 dB

- Sounds sources spatially separated
- Sound sources incompletely externalised

Auditory Neuroscience Laboratory,
University of Sydney
Results summary

Comparison between ITD + ILD condition and VAS condition

Target level with respect to Maskers (dB)

Conditions

Comparison of SRTs:
Full-band VAS and 5 kHz Low-passed VAS

Level of Target with respect to Maskers (dB)

Subjects
Conclusions

- Percept of spatialisation supports significant speech unmasking and improved speech intelligibility.
  - Is this mediated through selective attention and streaming?
- Speech energy >5 kHz plays a significant role in multi-talker spatial listening
  - Implications for multi-channel communications, virtual reality displays, hearing aid design and bio-inspired ASR.