Cocktail Party Phenomenon

(Bronkhorst (2000))

Madhusudana Shashanka

Auditory Neuroscience Laboratory
Department of Cognitive and Neural Systems
Boston University
Overview

Main focus
- Intelligibility of speech presented against a background of competing speech.

Organization
- Interfering speech sounds.
- Effect of interfering sounds during monaural or diotic presentation.
- Effect of a single interfering sound on binaural speech intelligibility.
- Binaural speech intelligibility for multiple interfering sounds.
- Effects of hearing impairment and the use of hearing aids.
- Conclusions.
Overview

- Interfering speech sounds.
- Monaural or diotic presentation.
- Binaural speech intelligibility – single interfering sound.
- Binaural speech intelligibility – multiple interfering sounds.
- Hearing impairment and hearing aids.
- Conclusions.
Interfering speech sounds

- Speech – broadband sound.
- Freq spectrum varies as a function of time.
- Useful spectral measure → long-term average spectrum.
  - Constant across languages when data averaged over several speakers.
  - Shape relatively constant, level quite variable.
- Long-term average spectrum level subject to
  - Interindividual differences in what one considers as conversational levels (up to $10\ dB$).
  - Distance - for *nearby* speakers, $6\ dB$ attenuation per doubling of distance.
  - Orientation of the speaker.
  - Lombard effect.

Bronkhorst: Cocktail party phenomenon – p.4/24
Interfering speech sounds

- Analysis based on signal-to-noise (S/N) ratio.
- When the number of speakers is large
  - babble $\sim$ long-term average speech spectrum.
- In case of few nearby speakers
  - Level fluctuations important.
  - Analyze modulations of speech envelope in different freq bands.
Overview

- Interfering speech sounds.
- Monaural or diotic presentation.
- Binaural speech intelligibility – single interfering sound.
- Binaural speech intelligibility – multiple interfering sounds.
- Hearing impairment and hearing aids.
- Conclusions.
Measures of speech intelligibility

- Articulation Index (AI), Speech Intelligibility Index (SII).

Important assumptions
- Contribution to AI in a narrow freq band $\propto$ S/N in that band.
- Contributions of non-overlapping bands can be combined by calculating a weighted sum.

- In the presence of time-domain distortions (reverberation, peak-clipping)
  - AI cannot be applied directly.
  - Speech Transmission Index (STI) - adds measure of time-domain effects based on modulation transfer function.
Voice vs noise

- At the same level, interfering voice a much less effective masker than noise.
- Release of masking when
  - 8-talker babble replaced by speech – \(10 \, dB\) (Miller).
  - White noise mixed with two talkers replaced by speech – \(4 \, dB\) (Carhart et al.).
  - Speech-spectrum noise replaced by speech (normal or time-reversed) – \(7 \, dB\) (Duquesnoy).
  - Speech-spectrum noise replaced by speech – \(6 - 8 \, dB\) (Festen and Plomp).
  - Speech-spectrum noise replaced by noise modulated like speech – \(4 - 6 \, dB\) (Festen and Plomp).
  - Noise replaced by speech – \(8 \, dB\) (Peters et al.).
Excess masking

- Segregation difficult if target and interfering voices are alike.
  - Male voices for both target and interferer - SRT of about 0 dB (Stubbs and Summerfield, Drullman and Bronkhorst).
  - Time-reversed speech of the target speaker as masker - SRT increase of 7 – 11 dB (Festen and Plomp).
  - CMR has only a minor influence - 1.3 dB on average (Festen).
  - “Listening in the gaps” the dominant effect.

- Unresolved issues:
  - Single measure of voice similarity.
  - modeling “listening in the gaps”.

Listening in the gaps

- Depends on the properties of both interferer and the target.
- Interrupted noise instead of continuous noise as masker
  - Release of masking for repetition rates between $1$ and $100$ Hz, with a maximum around $10$ Hz (Miller and Licklider).
- For a quantitative explanation, answer –
  - How should the AI be integrated as a function of time?
  - How should context effects be quantified?
- Bronkhorst and Houtgast - average modulation reduction $m$
  \[ m = \frac{1}{1 + I_N/I_S} \] instead of averaging $I_N$.
- Bronkhorst et al. - interrupted speech → phoneme intelligibility → syllable intelligibility (assumes exp decay of useful info within the duration of a phoneme).
Overview

- Interfering speech sounds.
- Monaural or diotic presentation.
- Binaural speech intelligibility – single interfering sound.
- Binaural speech intelligibility – multiple interfering sounds.
- Hearing impairment and hearing aids.
- Conclusions.
Spatial separation

- Release of masking up to 10 dB when speech and noise sources are spatially separated (Dirks and Wilson, Hirsh).
- Binaural gain very small when sources coincide (∼ 1 dB).
- Two contributions for the release of masking
  - Head shadow (high frequencies).
  - Interaural time delays (low and mid frequencies).
- The above gains are not additive for unprocessed noise (Bronkhorst and Plomp).
- “Monaural” thresholds affected by incomplete occlusion of the non-test ear (MacKeith and Coles).
- Binaural unmasking not sensitive to interaural differences in presentation level (Bronkhorst and Plomp).
Contd...

- Prediction of head shadow effects fairly straightforward.
- Modeling binaural gain due to ITD more difficult.
- Vom Hövel – AI model + modified EC model.
- Zurek – AI calculation + simplified Colburn’s auditory-nerve-based model.
  - Trends in data predicted rather well.
  - Individual differences can be quite large (up to $5 \text{ dB}$).
- Zurek’s predictions
  - Calculation of the optimal head orientation given the positions of speech and masker sources.
  - Head rotations (within $\pm 45^\circ$) results in an advantage of $3 \text{ dB}$.
Realistic Environments

- Reverberation and background noise – performance deterioration.
  - Target speech less intelligible.
  - Smaller binaural release from masking.
  - Less gain from masker fluctuations.

- Binaural gain in the presence of reverberation, even in the absence of competing sound sources.

- Release from masking due to spatial separation of sources significantly low (4 dB - Bronkhorst and Plomp, Koehnke and Besig).

- Farther noise source → reduced release from masking due to masker fluctuations (Bronkhorst and Plomp).
Overview

- Interfering speech sounds.
- Monaural or diotic presentation.
- Binaural speech intelligibility – single interfering sound.
- Binaural speech intelligibility – multiple interfering sounds.
- Hearing impairment and hearing aids.
- Conclusions.
Multiple interfering sounds

- Different maskers presented separately to the two ears (Pollack and Pickett)
  - No interaural time or level differences.
  - Binaural release from masking – 13 to 5 dB when number of voices per ear increased from one to seven.

- Interaural time and phase differences applied to multiple maskers (Carhart et al.)
  - Binaural release from masking does not decrease as a function of the number of interfering sounds.
  - No relation between unmasking and lateralization.
Sources and spatial distribution

Speech intelligibility determined by
- Number of sources.
- Spatial distribution.
- Temporal envelope of the interfering sounds.

Release of masking due to spatial separation decreases with increasing number of sources.
- Due to reduction of the head shadow component (Bronkhorst and Plomp).
- Difference between better and worse-ear performance decreases when maskers are added.
- Difference between binaural and better-ear performance remains the same.

Effect of configuration is always larger than that of adding/removing one source (Hawley et al.).
Sources and spatial distribution

Release from masking $R$ due to spatial separation (target speech presented frontally)

$$R = C\left[ \alpha \left(1 - \frac{1}{N} \sum_i \cos \theta_i \right) + \beta \frac{1}{N} \left| \sum_i \sin \theta_i \right| \right].$$

Assumptions - sources in the horizontal plane, maskers have same long-term average level.

Constant $C$ - accounts for differences due to the choice of method and speech material.

$\sin$ term - quantifies the degree of asymmetry of the configuration (binaural gain).

$\cos$ term - quantifies proximity to the target source (increased masking).

Target at azimuth $\phi$ - use $(\theta - \phi)$ in the $\cos$ term (insufficient data for validation).
Divided attention tasks

- Main effects - number and spatial configuration of sources.
- Spatial configuration seems to have a larger effect (Yost et al.).
- Binaural gain increases as a function of the number of sources – reasons?
  - Ceiling effects.
  - Head shadow effects not included for the monaural condition.
- Speech-recognition data (Drullman and Bronkhorst) – large effects in the response times.
Overview

- Interfering speech sounds.
- Monaural or diotic presentation.
- Binaural speech intelligibility – single interfering sound.
- Binaural speech intelligibility – multiple interfering sounds.
- Hearing impairment and hearing aids.
- Conclusions.
Hearing impairment

- Hearing loss for speech in noise.
  - Less benefit from fluctuations in the interfering sound when compared to normal hearing persons (Carhart and Tillman).
  - Gain of only $0-2 \, \text{dB}$ (instead of $7 \, \text{dB}$) when noise replaced by interfering speech.
  - Release from masking due to spatial separation affected (only head-shadow component).

- Hearing aids - benefit can be expected if (Plomp)
  - Signals are presented within the available dynamic range.
  - Negative effects of upward spread of masking should be prevented.
  - Promising results - microphone arrays and processing of binaural signals.
Overview

- Interfering speech sounds.
- Monaural or diotic presentation.
- Binaural speech intelligibility – single interfering sound.
- Binaural speech intelligibility – multiple interfering sounds.
- Hearing impairment and hearing aids.
- Conclusions.
Topics for further study

- Effects of hearing impairment and benefits of hearing aids.
- Excess masking, related to voice segregation
  - Is the interfering effect of certain voices more than what could be expected given their frequency spectrum and temporal modulations?
- Speech intelligibility in realistic circumstances
  - Not much known about noise levels that actually occur.
  - Research on binaural speech perception in the presence of reverberation scarce.
- Influence of attention
  - Would performance be better for selective than for divided attention tasks?