Talker identification in three types of background noise

Alina Razak, Emily J. Thurston, Lauren E. Gustainis, Gerald Kidd, Jr., Jayaganesh Swaminathan, & Tyler K. Perrachione

Department of Speech, Language, & Hearing Sciences, Boston University

BOSTON UNIVERSITY

Summary

- Perceiving speech in the presence of background noise (i.e., masking) requires listeners to distinguish and attend to one talker from among competing talkers.
- Young adults with normal hearing learned to identify five talkers by the sound of their voice in quiet.
- Talker identification accuracy was tested in quiet and masking by speech-shaped noise, multi-talker babble, and a single competing talker.
- Performance was best in quiet and fell as a function of SNR for all masker types.
- The change in accuracy as a function of SNR was greatest for multi-talker babble.
- Trained sentences facilitated identification compared to untrained sentences.
- Individual variability in performance appeared unrelated to listeners' pure-tone hearing levels, speech perception in noise, or phonological awareness/memory.

Methods

Participants

Native English-speaking young adults (N = 30; 25 female, 5 male; age 18-30) with a self-reported history free from speech, language, and hearing problems.

Target Talkers and Sentences

Participants learned to identify 5 female native speakers of American English from their recordings of 30 “Harvard Sentences” (IEEE, 1969). Listeners trained on 15 of the sentences and tested on all 30. Training/testing sentences were counter-balanced across participants.

Masking Noise and Speech

- Speech-shaped noise was extracted from track 69 of the Quick Speech-in-Noise Test (QuickSIN; Killion *et al.*, 2004).
- Four-talker babble was extracted from tracks 24-35 of the QuickSIN.
- Single competing talkers (N = 4, female native speakers of American English) recorded 26 additional “Harvard Sentences”.
- Combined RMS amplitude of targets and maskers was held constant at 68dB across the different test SNRs (Inf, +5dB, 0dB, -5dB).

Procedure

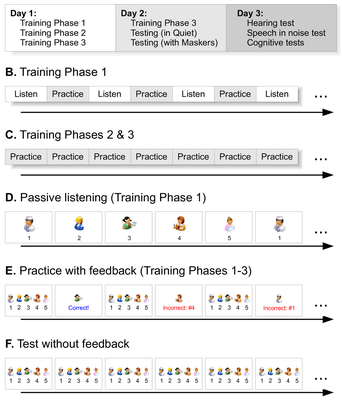
(A) Three visits for talker identification training/testing and cognitive / audiological assessment.

(B) Phase 1: Passive familiarization of talkers blocked by sentence **(D)**, followed by active practice with feedback **(E)** (Day 1).

(C) Phase 2: Further active practice with feedback blocked by sentence **(E)** (Day 1); Phase 3: Further active practice with feedback with random sentence order **(E)** (Day 1 & 2).

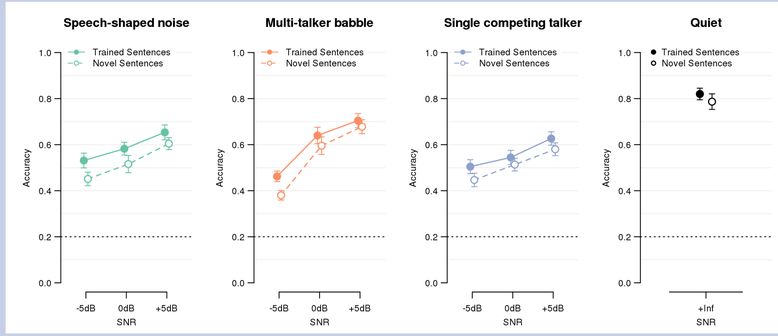
(F) Talker identification test with no feedback, first in quiet, and then in the three types of masking noise (SSN, MTB, SCT; counterbalanced) (Day 2).

Day 3: Pure-tone hearing (Carhart *et al.*, 1959), speech-in-noise (QuickSIN), and cognitive tests (CTOPP; Wagner *et al.*, 2013).



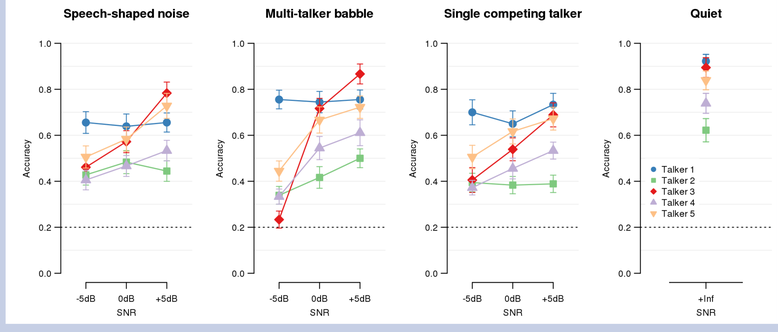
Results: Masking Conditions

Talker identification accuracy as a function of SNR across noise types



- Marginal effect of masker type such that overall talker identification accuracy in background babble (57.7%) was somewhat better than in speech-shaped noise (55.6%) or a single competing talker (53.6%) [$F_{2,58}$ = 2.73, p = 0.073].
- Significant effect of SNR such that overall talker identification accuracy differed as a function of SNR [$F_{1,29}$ = 167.86, p = 1.38×10^{-13}].
- Participants were most accurate in the most favorable SNR, and accuracy fell as SNR decreased (+5dB: 64%; 0dB: 57%; -5dB: 46%).
- Facilitatory effect of SNR varied across noise type [condition by SNR; $F_{2,58}$ = 9.52, p = 2.65×10^{-4}] such that multi-talker babble had the steepest slope for talker identification accuracy.
- Overall talker identification accuracy for trained sentences (60.7%) was significantly better than for novel sentences (55.5%) [$F_{1,29}$ = 28.05, p = 1.1×10^{-5}].
- Facilitatory effect of trained versus novel sentences did not differ across masker type [$F_{2,58}$ = 0.37, p = 0.69] nor SNR [$F_{1,29}$ = 1.98, p = 0.17].
- No three-way interaction between condition, SNR, and sentence familiarity was observed [$F_{2,58}$ = 0.36, p = 0.70].

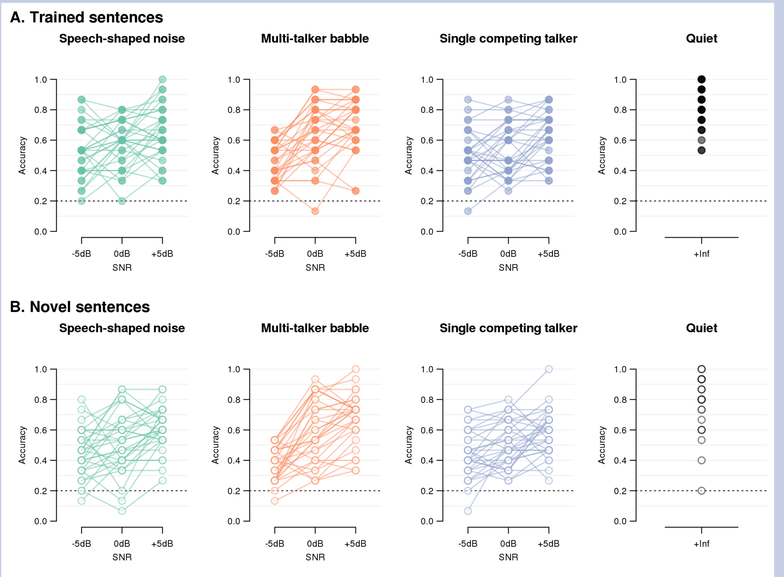
Talker identification accuracy depending on target speaker



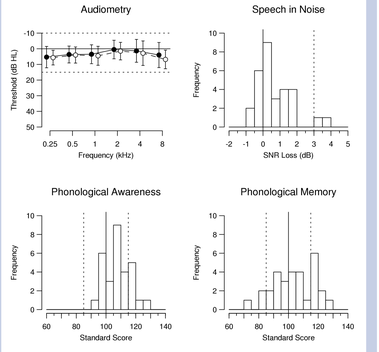
- There was substantial variability in talker identification accuracy across speakers.
- While some remained unaffected by background noise (Talker 1), others varied across both noise type and SNR (most notably, Talker 3).
- The most identifiable talker at +5dB was poorly identified at -5dB SNR (Talker 3).

Results: Individual Variability

Variability in talker identification accuracy across participants



Supplemental assessments



- Individual performance varied considerably across participants.
- Correlation between talker identification accuracy in quiet and in masking was significant (SSN: r = 0.78, p < 4×10^{-5}; MTB: r = 0.81, p < 5×10^{-8} ; SCT: r = 0.67, p < 5×10^{-5}).
- After controlling for accuracy in quiet, no significant correlations were found between talker identification accuracy with masking and the supplemental assessments ($|r|$ = 0.03 – 0.28).
- Decrement in accuracy in babble from +5dB to -5dB was marginally correlated with the phonological awareness (r = 0.44, p < 0.02) and phonological memory (r = 0.45, p < 0.02) subtests of the CTOPP.

References

Carhart *et al.* (1959). *J. Speech and Hearing Disorders.* 24, 330-345.

Institute of Electrical and Electronics Engineers. (1969). *IEEE Transactions on Audio and Electroacoustics.* 17, 225-246.

Killion *et al.* (2004). *J. Acoust. Soc. Am.* 116, 2395.

Wagner *et al.* (2003). *Comprehensive Test of Phonological Processing* – 2, Pro-Ed.

Acknowledgments

We thank Lorraine Delhorne, Cassandra Chan, Jennifer Golditch, Andrea Chang, Deirdre McLaughlin, Virginia Best, Christine Mason, & Elin Roverud for their assistance. This study was supported by the NIDCD of the NIH under award number R03DC014045 (to TP) and DC004545 (to GK).

BU Communication Neuroscience Research Laboratory

cnrlab@bu.edu

http://sites.bu.edu/cnrlab/