Listener strategy and performance in linguistic and non-linguistic auditory divided attention tasks

Daniel R McCloy · Lindsey R Kishline · Adrian KC Lee

Department of Speech and Hearing Sciences & Institute for Learning and Brain Sciences, University of Washington


DRM


LRK


AKCL

BACKGROUND

- Vision research supports a "spotlight" model of selective and divided attention [1, 2].
- Previous auditory research has not conclusively established which model is most appropriate for audition [3, 4, 5].

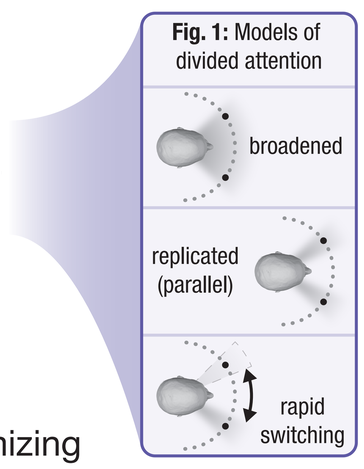


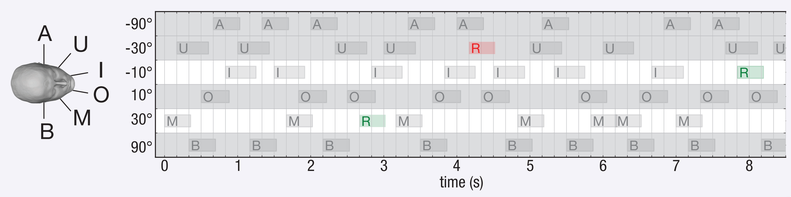
Fig. 1: Models of divided attention

OUR RESEARCH

- A novel auditory paradigm with 4-6 temporally interleaved streams defined by spatial location (simulated by convolution with HRTFs [6]), minimizing spectrotemporal (energetic) masking.
- A series of oddball detection experiments using spoken alphabet letters or monosyllabic words, with manipulations of target spatial separation and linguistic complexity of task.
- **Questions:** which model(s) are accurate representations of listener strategy? How does task type influence listener behavior? How do the cortical circuits for attention and speech processing interact?

EXPERIMENT 1: NUMBER OF ATTENDED STREAMS

Fig. 2: Sample trial structure (first 8 waves). As shown, to-be-attended streams (white backgrounds) are spatially separated. Small grey rectangles are letter durations; targets are green, foils are red.



Methods summary: 6 streams of spoken letters (talker *famd0*, ISOLET corpus) at ±10°, ±30°, ±90° with fixed relationship between letter and angle; targets always "R"; 21 reps. per stream; 1-6 streams visually cued as to-be-attended.

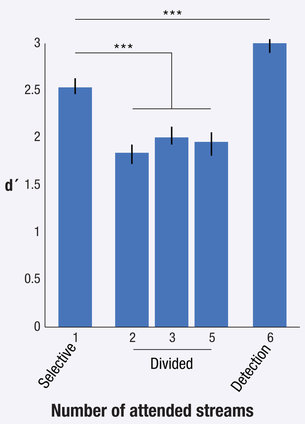
Results: Fig. 3: sensitivity (d′) higest for *detection* (attend all 6 streams); intermediate for *selective attention* (attend 1 stream); lowest for *divided attention* (attend 2/3/5 streams).

In divided attention condition, attended streams adjacent > separated (not shown).

Fig 4: false alarms to streams spatially interposed between attended streams (e.g., the "O" stream in Fig. 2) are more common than FAs to outside streams (the "U" stream).

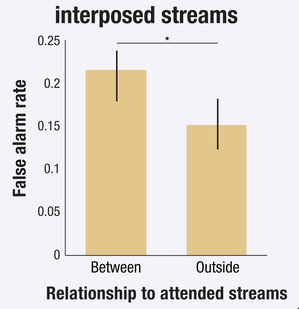
Take-away: it is hard to monitor spatially separated auditory targets while ignoring spatially interposed streams.

Fig. 3: d′ vs. attended streams



n=16

Fig. 4: False alarm rate for interposed streams



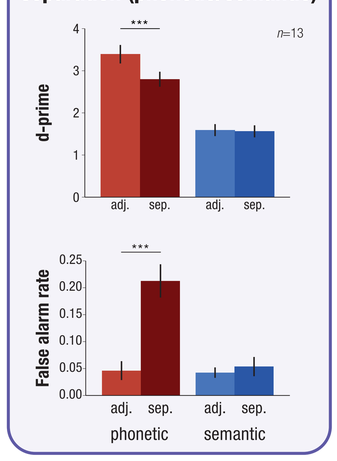
EXPERIMENT 2: PHONETIC vs SEMANTIC DEVIANTS

Methods summary: 4 monosyllabic word streams at ±15°, ±60° with fixed relationship between word/category and angle; 12 reps. per stream; 1-2 streams visually cued as to-be-attended. *Phonetic*: single "base" word per stream; target is any other word. *Semantic*: each stream has 6 related words; targets are semantic deviants.

Results: Replicated adjacent/separated result from Exps. 1 & 2, but only in phonetic task. Effect strongly driven by false alarm rate.

Take-away: the challenge of ignoring interposed streams may depend on task type or difficulty (no elevated false alarm rate to interposed streams in semantic task).

Fig. 5: d′ & FA rate vs. spatial separation (phonetic/semantic)



EXPERIMENT 3: LINGUISTIC TASK COMPLEXITY

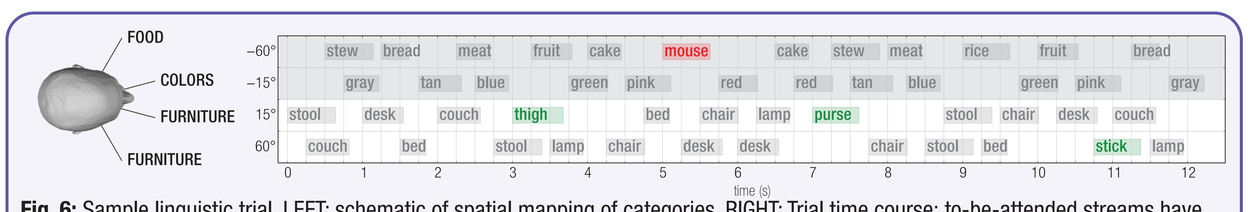


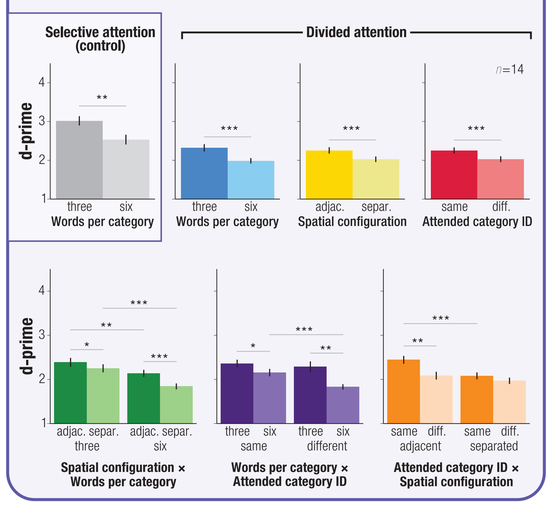
Fig. 6: Sample linguistic trial. LEFT: schematic of spatial mapping of categories. RIGHT: Trial time course; to-be-attended streams have white backgrounds, ignored streams have grey backgrounds; small rectangles are actual word durations; targets are green, foils are red.

Methods summary: 4 word streams, 2 streams attended. Manipulation: congruence of attended streams (same vs. different category) and category size (3 vs. 6 words).

Results: Fig 7: d′ improves when fewer words per category (blue), attended streams are adjacent (yellow), attended streams are same semantic category (red).

Interactions: more words per category → harder when spatially separated (green) or when categories incongruent (purple). Benefit when spatially adjacent categories are congruent (orange).

Fig. 7: d′ main effects & interactions: category size, congruence, and spatial adjacency



GENERAL DISCUSSION

- Listeners may have multiple strategies; models not mutually exclusive.
- **Future work:** Assess listener effort with pupillometry, segregate cognitive load due to linguistic / non-linguistic aspects of the task; M/EEG neuroimaging to identify convergence of attentional / linguistic cortical circuits.

REFERENCES

[1] Eriksen CW & Hoffman JE (1972). Temporal and spatial characteristics of selective encoding from visual displays. Percept Psychophys, 12(2), 201–204. doi:10.3758/BF03212870

[2] Eriksen CW & St James JD (1986). Visual attention within and around the field of focal attention: A zoom lens model. Percept Psychophys, 40(4), 225–240. doi:10.3758/BF03211502

[3] Hafter E, Xia J, Kalluri S, Poggesi R, Hansen C & Whiteford K (2013). Attentional switching when listeners respond to semantic meaning expressed by multiple talkers. POMA 19, 050077. doi:10.1121/1.4801413

[4] Best V, Gallun F, Ihlefeld A, & Shinn-Cunningham B (2006). The influence of spatial separation on divided listening. J Acoust Soc Am 120(3), 1506–1516. doi:10.1121/1.2234849

[5] Rivenez M, Darwin C & Guillaume A (2006). Processing unattended speech. J Acoust Soc Am 119(6), 4027–4040. doi:10.1121/1.2190162

[6] Shinn-Cunningham B, Kopco N & Martin T (2005). Localizing nearby sound sources in a classroom: Binaural room impulse responses. J Acoust Soc Am 117(5), 3100–3115. doi:10.1121/1.1872572

ACKNOWLEDGMENTS

This research was funded by R01–DC013260 to Adrian KC Lee, and T32–DC000033 to the Department of Speech and Hearing Sciences, University of Washington. Special thanks to Eric Larson and Ross Maddox for helpful discussions and feedback.