The multimodal nature of spoken word processing in the visual world: Testing the predictions of alternative models of multimodal integration

Alastair C. Smith (alastair.smith@mpi.nl)
Max Planck Institute for Psycholinguistics, Netherlands

Padraic Monaghan (p.monaghan@lancaster.ac.uk)
Department of Psychology, Lancaster University, United Kingdom

Falk Huettig (falk.huettig@mpi.nl)
Max Planck Institute for Psycholinguistics, Netherlands;
Donders Institute for Brain, Cognition and Behaviour, Radboud University, Netherlands

Ambiguity in natural language is ubiquitous (Piantadosi, Tily & Gibson, 2012), yet spoken communication is effective due to integration of information carried in the speech signal with information available in the surrounding multimodal landscape. However, current cognitive models of spoken word recognition and comprehension are underspecified with respect to when and how multimodal information interacts in the cognitive system.

Within this study we investigate this issue by comparing two computational models both of which frame spoken word recognition and speech comprehension in terms of multimodal constraint satisfaction. Both models permit the integration of concurrent information within linguistic and non-linguistic processing streams, however their architectures differ critically in the level at which multimodal information interacts. We compare the predictions of the Multimodal Integration Model (MIM) of language processing (Smith, Monaghan & Huettig, 2014), which like 'hub and spoke' models of semantic processing (Plaut, 2002; Rogers et al., 2004; Dilkina, McClelland, & Plaut, 2008), implements full interactivity between modalities, to a model in which interaction between modalities is restricted to lexical representations which we represent by an extended multimodal version of the TRACE model of spoken word recognition (McClelland & Elman, 1986).

Language mediated visual attention requires visual and linguistic information integration and has thus been used to examine properties of the architecture supporting multimodal processing during spoken language comprehension (Huettig, Rommers & Meyer, 2011). We generate predictions from these alternative models for the influence of visual, semantic and phonological rhyme similarity on language mediated visual attention that are then tested in two visual world experiments.

Our results demonstrate that previous visual world data sets involving phonological onset similarity are compatible with both models, whereas our novel experimental data on rhyme similarity is able to distinguish between competing architectures. The fully interactive MIM system correctly predicts a greater influence of visual and semantic information relative to phonological rhyme information on gaze behaviour, while by contrast a system that restricts multimodal interaction to the lexical level overestimates the influence of phonological rhyme, predicting stronger effects of phonological rhyme relative to semantic and visual information, thereby providing an upper limit for when information interacts in multimodal tasks.

We discuss the continued under-specification of the representational structures and cognitive architecture supporting multimodal language processing and how novel properties of the deep learning approach offer potential for new insight on these issues that are fundamental to our understanding of language processing.

References

Dilkina, K., McClelland, J. L., & Plaut, D. C. (2008). A single system account of semantic and lexical deficits in five semantic dementia patients. *Cognitive Neuropsychology, 25*, 136–164.

Huettig, F., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica*, *137(2)*, 151-171.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. Cognitive Psychology, 18(1), 1-86.

Piantadosi, S. T., Tily, H., & Gibson, E. (2012). The communicative function of ambiguity in language. *Cognition*, *122(3)*, 280-291.

Plaut, D. C. (2002). Graded modality-specific specialisation in semantics: a computational account of optic aphasia. *Cognitive Neuropsychology*, *19*, 603–639.

Rogers, T. T., Lambon Ralph, M. A., Garrard, P., Bozeat, S., McClelland, J. L., Hodges, J. R., & Patterson, K. (2004). Structure and deterioration of semantic memory: a neuropsychological and computational investigation. *Psychological review*, *111(1)*, 205.

Smith, A. C., Monaghan, P., & Huettig, F. (2014). Literacy effects on language and vision: Emergent effects from an amodal shared resource (ASR) computational model. *Cognitive Psychology*, *75*, 28-54.