

# RESEARCH DIRECTIONS

## Listen carefully, one syllable at a time

Dr Jason Shaw, of the MARCS Institute, has been awarded funding to investigate the cognitive mechanisms that underpin the human ability to recognise both words and talkers in speech. The project, which is supported by an Australian Research Council Discovery Early Career Researcher Award, will produce a blueprint for technologies that integrate speech recognition with talker recognition.

'Human speech involves individual talkers with individual speech characteristics. The background of a talker, their physical properties and their emotional state all affect how words are pronounced,' says Dr Shaw. 'Humans can typically recognize a familiar talker and learn properties of new talkers. By seven months of age, for example, normally developing infants are able to disentangle linguistic structure from talker identity.'

Understanding how humans accomplish this so easily is a major challenge of modern cognitive science. Attempts to develop software applications which automatically recognize words in human speech, and keep track of variations in speech introduced by different talkers, have found this a formidable obstacle. Humans, meanwhile, effortlessly and automatically parse the speech stream into words.

A growing body of research indicates that *linguistic* (that which identifies words) and *indexical* (that which identifies the talker) information interacts. Words are recognised faster in speech produced by a familiar talker; talker identification is made much more difficult when the talker speaks a different language to the listener. Dr Shaw's project proposes that talker distinctness is processed through the linguistic frame of the syllable. There is now considerable evidence that the syllable affects the relative timing between articulatory organs, such as the lips, tongue, jaw and soft palate. His research includes implementing that idea in a



computational model of syllable structure and testing it through a series of perceptual studies. Relatively little research has investigated whether the individual variation in timing can be used by listeners to discriminate talkers. Dr Shaw aims to quantify variation in speech timing in a large multi-speaker body of Australian English; determine the extent to which listeners make use of talker-specific timing patterns to recognize talkers; develop a computational model integrating indexical and linguistic aspects of speech timing; and test model predictions for talker distinctiveness against human performance in discriminating talkers.

The development of a new theory of what is shared across speakers and what sets them apart has implications for verifying "ear witness" testimony in courts and the development of dual automatic speech recognition systems that identify both words and talkers simultaneously from running speech.

**Project Title:** How we know who is talking: talker-distinctiveness in speech timing

**Funding has been set at:** \$375,000

**Contact Details:** [j.shaw@uws.edu.au](mailto:j.shaw@uws.edu.au)

URL - <http://marcs.uws.edu.au/>

**January 2014**