

Auditory kernels – learning acoustic building blocks from speech

Dimme de Groot¹, Odette Scharenborg¹, Jorge Martinez¹

¹Technische Universiteit Delft

Unlike linguistic or lexical units such as phonemes or syllables, there are no obvious “information-carrying” acoustic units directly observable in the speech waveform. According to the efficient coding theorem, however, the auditory system should encode incoming sensory information as compactly as possible. Smith and Lewicki (2006) demonstrated that, when learning sparse acoustic building blocks—referred to here as auditory kernels—from speech, the resulting kernels closely resemble reverse-correlation (revcor) filters measured in cat auditory systems. In this work, we extend their analysis to a large cross-linguistic dataset comprising speech of 102 languages. We learn the auditory kernels from each language and examine their statistical properties. We find that the kernels learned on the different languages have a remarkably similar spectral centroid-spread relationship. We also find that, irrespective of language, around 10 kernels are used to represent the content below 500 Hz. These results suggest that this representation might be universal and encourage further research.