

Improving on the state-of-the-art: integrating linguistic-phonetic and automatic methods in forensic voice comparison

In forensic voice comparison (FVC) casework, an expert is typically instructed to analyse the speech in recordings of a known suspect and an unknown offender in order to help the court decide whether they contain the voices of the same or different speaker(s). Broadly speaking there are two methods of analysis commonly applied in FVC work: (i) linguistic-phonetic, involving a combination of auditory and acoustic analysis of linguistic elements of the speech signal (e.g. vowel formants, f_0 , articulation/speech rate); and (ii) automatic approaches, involving (typically, but not always) holistic analysis of the entire speech signal using features such as Mel-frequency cepstral coefficients (MFCCs). However, these approaches have largely developed in parallel to each other with almost no work (with the exception of Gonzalez-Rodriguez et al., 2014) considering the relative contribution of the two methods to speaker characterisation.

The *Voice and Identity* project at York is the first large-scale study to integrate linguistic-phonetic and automatic approaches to examine how speaker identity is encoded in the speech signal and how the best elements of the two methods may be combined to optimise the performance of ASR systems. In this paper I will discuss the underlying principles, limitations and performance of state-of-the-art ASR systems. I will also present results which show how the performance of such ASR systems may be improved with the inclusion of a relatively small amount of linguistic-phonetic data. Finally, I will consider the future of FVC methods and the contribution of the *Voice and Identity* project to the field.