

2pSCa4. Detecting voice disguise from speech variability: Analysis of three glottal and vocal tract measures. Talal B. Amin (Elec. and Electron. Eng., Nanyang Technolog. Univ., Singapore, Singapore), James S. German (Humanities and Social Sci., Nanyang Technolog. Univ., HSS-03-46, Singapore, Singapore, jsgerman@ntu.edu.sg), and Pina Marziliano (Elec. and Electron. Eng., Nanyang Technolog. Univ., Singapore, Singapore)

The deliberate attempt by speakers to conceal their identity (voice disguise) presents a challenge for forensics and for automated speaker identification systems. Using a database of natural and disguised voices of three professional voice impersonators, we build on earlier findings (Amin et al., 2012) by exploring how certain glottal and vocal tract measures, including fundamental frequency (f_0), glottal timing (Open Quotient), and vowel formants, are exploited to create novel voice identities. Specifically, we explored whether the amount and type of variation exhibited by impersonators can be used to develop a metric for distinguishing natural from disguised voices. As expected, variation in f_0 and Open Quotient was speaker dependent, and corresponded closely to social attributes (i.e., gender/age) of the voice identities involved. In a novel finding, the effects of voice identity on vowel formants were highly dependent on vowel category, and could not be readily characterized as global modifications to the vowel space (Bradlow et al., 1996). We therefore developed a no-reference objective metric for voice disguise that treats formant variability on a vowel-by-vowel basis. This metric consistently assigned high rankings to natural voices (3.3/27 on average). This correlated closely with the subjective disguisedness ratings of 18 naïve listeners, even outperforming them slightly.