Spoofed Samples: another thing to listen out for?

Georgina Brown[1,2]*, Lois Fairclough*[1] *and Christin Kirchhübel*[2]

[1]*Department of Linguistics and English Language, Lancaster University, UK*
[2]*Soundscape Voice Evidence, Lancaster, UK*

{g.brown5|l.fairclough}@lancaster.ac.uk, ck@soundscapevoice.com

"Spoofing" has been raised as a very real risk in the context of automatic speaker verification systems (Evans et. al., 2013). In spoofing attacks, speech samples are submitted to a speaker verification system with the intention of "tricking" the system into falsely accepting the sample as belonging to a specific speaker. Understandably, spoofing attacks are a growing concern among certain sectors in particular (such as the financial sector), where voice, as a "biometric", is increasingly being used as a mechanism to access accounts. There are four key spoofing methods: 1) impersonation; 2) replay; 3) speech synthesis; 4) voice conversion (Wu et. al., 2015a). *Impersonation* is perhaps the most intuitive, where it involves one human modifying their own voice to sound more like the voice of the "target" speaker. *Replay* refers to replaying a previously captured recording of the "target" speaker producing the specified utterance (or "passphrase") to a system. *Speech synthesis* refers to the technologies used to produce synthetic speech that sounds like a "target" sample, while *voice conversion* refers to technologies used to modify a speech sample to sound more like someone or something else (i.e. the "target").

Efforts to identify solutions to combat spoofing attacks have commenced within the speech technology community. The creation of the *ASVSpoof Challenge* (Wu et. al., 2015b) has enabled the international research community to pre-emptively innovate and advance countermeasures. The ASVSpoof challenges have become a regular event, taking place every two years. For these challenges, a team of researchers compile a database of thousands of short speech samples, based on read sentences. These large datasets allow other researchers to participate in the challenge where they can test their speaker verification systems on these speech samples (to determine how much of a threat specific spoofing techniques are), as well as to test new methods that aim to detect or counteract spoofing attacks. Another property of the ASVSpoof datasets is that the spoofed samples are produced by a wide range of spoofing techniques. In the 2015 challenge, the datasets contained spoofed samples produced by 10 different speech synthesis and voice conversion techniques, while this number increased to 17 for the 2019 challenge. Given the speed at which speech technologies are developing, it is reassuring to know that anti-spoofing research is now taking place in parallel.

While the central focus of anti-spoofing countermeasures is very much on automatic speaker verification systems, the current work starts to contemplate the potential of spoofed speech samples occurring in forensic casework. Forensic speech practitioners already have to occasionally contend with some form of "spoofing" in the form of voice disguise, but it seems sensible to extend our knowledge to account for more technologically-derived forms. Rather than assuming that spoofed speech samples would be detectable to an expert forensic phonetician, the authors of this work have chosen to test this assertion. Taking the datasets used to develop and evaluate anti-spoofing technologies, the current paper reports on how one experienced forensic phonetician performed in a simple test that asked for spoofing evaluations of 300 speech samples (some were spoofed samples, some were genuine human speech samples). Within this set of 300 speech samples, there are 150 samples from the ASVSpoof 2015 Challenge (Wu et. al., 2015b), and 150 from ASVSpoof 2019 Challenge

(Todisco et. al., 2019). This was in an effort to track any change in the quality (or risk) of spoofing attacks over time. We also selected the spoofing techniques that were reported to be particularly problematic for automatic technologies (Wu et. al., 2015b; Todisco et. al., 2019). We included spoofed samples produced by the most challenging voice conversion technique and the most challenging speech synthesis technique from each of the two ASVSpoof Challenge datasets. Out of the selection of spoofing techniques that have been included in our test set, the "most successful" one brought about Equal Error Rate of 57.73% from the automatic speaker verification system used in Todisco et. al. (2019).

Not only do we report on the test results, but we also impart qualitative observations on reflection of this test. We also propose it as a valuable training exercise for forensic speech analysts, and offer the opportunity to others in the community to take the test.

References

Evans, N., Kinnunen, T. and Yamagishi, J. (2013). Spoofing and Countermeasures for Automatic Speaker Verification. *Proceedings of Interspeech*. Lyon, France. 925-929.

Todisco. M., Wang, X., Vestman, V., Sahidullah, M., Delgado, H., Nautsch, A., Yamagishi, J., Evans, N., Kinnunen, T. and Lee, K.A. (2019). ASVspoof 2019: Future Horizons in Spoofed and Fake Audio Detection. *Proceedings of Interspeech*. Graz, Austria. 1008-1012.

Wu, Zhizheng., Evans, N., Kinnunen, T., Yamagishi, J., Alegre, F. and Li, H. (2015a). *Speech Communication*. 66. 130-153.

Wu, Zhizhen, Kinnunen, T., Evans, N., Yamagishi, J., Hanilci, C., Sahidullah, M. and Sizov, A. (2015b). ASVspoof 2015: the First Automatic Speaker Verification Spoofing and Countermeasures Challenge. *Proceedings of Interspeech*. Dresden, Germany. 2037-2041.