Chapter 5
Voice Liveness Detection for Medical Devices

Bin Hao
University of Louisiana at Lafayette, USA

Xiali Hei
University of Louisiana at Lafayette, USA

ABSTRACT

Many healthcare providers integrate biometric recognition/verification schemes into patient identification or other information security systems. While overcoming the disadvantages of using passwords, PINs, and tokens which may be forgotten, or stolen, biometric systems are susceptible to spoofing attacks, or presentation attacks. Liveness detection is an effective mechanism used to defeat a presentation attack. This chapter focuses on voice liveness detection in automatic speaker verification (ASV) systems. The authors explain the spoofing attacks to ASV systems comprising impersonation, voice conversion, speech synthesis, and replay and then present four types of liveness detection (anti-spoofing) methods used to mitigate ASV spoofing attacks: challenge-response-based methods, acoustic feature-based methods, hardware-based methods, and multi-modal biometric-based methods. This chapter analyzes the advantages and disadvantages of each kind of liveness detection method and proposes the possible application of voiceprint-based liveness detection schemes in the insulin pump system.

INTRODUCTION

According to the analysis report from Crystal Market Research (CMR), the global healthcare biometrics market will reach around $12 billion by 2025. This growth is due to the fact that more and more healthcare providers integrate biometric recognition/verification schemes to patient identification or other information security systems. Biometric identification systems establish the identity of the users based on their extracted physiological features including face, fingerprint, iris, retina, vein (finger or palm), palm geometry, etc. or behavioral features including voice, signature, keystroke dynamics, etc.

DOI: 10.4018/978-1-5225-7525-2.ch005

While overcoming the disadvantages of using passwords, PINs, and tokens which may be forgotten, or stolen, biometric systems are susceptible to spoofing attacks, or Presentation Attacks (PA) which outwit a biometric sensor by presenting a counterfeit biometric evidence of a legitimate user using methods such as artifact, mutilations, replay, etc. to achieve impersonation or concealment.

Liveness detection is an effective mechanism used to detect a presentation attack. Recently, liveness detection has become a hot research topic in fingerprint recognition, iris recognition, and automatic speaker verification (ASV) communities. Research results from LivDet Iris 2017, LivDet Fingerprint 2017, and ASVspoof 2017 show that liveness detection or presentation attacks detection (PAD) systems still need to make more advancements, especially when under unknown attacks.

This chapter mainly focuses on voice liveness detection in ASV systems. An AVS system extracts the vocal characteristics of an individual to establish the identity either by imposing the fixed vocabulary constraints (text-dependent) or in a dynamic way (text-independent) i.e. without imposition of vocabulary constraints on the individuals. Currently, ASV is mature technique ready for commercial application in user authentication. But it is confirmed that ASV is vulnerable to spoofing attacks which undermine users' confidence on it. ASV spoofing attacks comprise (Wu et al., 2015a): impersonation whereby an attacker attempts to mimic a target legitimate user's voice; voice conversion whereby an attacker resembles the speech of a target legitimate user using another user's speech; speech synthesis whereby an attacker using text-to-speech (TTS) technique generates intelligible, natural-sounding artificial speech with inputs of arbitrary text; and replay whereby an attacker tries to pass the authentication of ASV by providing a pre-recorded speech sample collected from genuine target legitimate user.

According to state-of-the-art research results, there are mainly four kinds of voice liveness detection methods as anti-spoofing Countermeasures (CM) to mitigate ASV spoofing attacks.

- **Challenge-Response-Based Methods:** Aley-Raz et al. (2013) proposed a speaker recognition system with liveness detection scheme. This mechanism requires users' explicit cooperation (provide a sentence selected randomly from a closed set of sentences for liveness detection) when the voice authentication is processing.
- **Acoustic-Feature-Based Methods:** The basic idea of acoustic-feature-based methods comes from the observation that design of spoofing countermeasures should focus on the search for discriminative features rather than the design of complex classifiers. Many features such as Mel Frequency Cepstral Coefficients (MFCCs) and Linear Prediction Cepstral Coefficients (LPCCs), have been used for training classifiers (Font et al., 2017). For backend classifiers, Gaussian Mixture Model (GMM), GMM with Universal Background Model (GMM-UBM), Support Vector Machine (SVM), Deep Neural Networks (DNN)-based classifiers are the most often choices.
- **Hardware-Based Methods:** Chen et al. (2017) proposed a robust smartphone based voice impersonation defense system that captures the magnetic field emitting from the loudspeaker and extracts special physical characteristics of the magnetic field to discriminate between a live human and a loudspeaker (impersonation attacker). Zhang et al. (2016) proposed a liveness detection scheme (VoiceLive) that uses phoneme sound localization, i.e., the time-difference-of-arrival (TDoA) changes to the two microphones equipped in a smartphone, to determine whether a passphrase is spoken by a live person or replayed by loudspeaker. Zhang et al. (2017) proposed VoiceGesture that leverages the response of human speech production system to external stimuli to implement liveness detection.