Chaitravi Reddy  
16010422274

***VULNERABILITY IN SPEAKER VERIFICATION – A***

***STUDY OF TECHNICAL IMPOSTOR TECHNIQUES***

We explore scenarios where impostors have knowledge of the system and access to target speech to deceive the system. By testing a worst-case scenario, the study aims to assess the upper limit of vulnerability in SV systems against intentional impostor.  
Hidden Markov Model (HMM) based speaker verification system is used as a reference to investigate imposture techniques. The experiments are conducted using a speaker verification database, where the intentional impostor has access to a portion of the database while the SV system is trained on a separate portion. The SV system uses whole word left-right HMMs with specific password sequences. The study focuses on two speakers from the database, a male and a female from the same dialect region and age group.

**Experiment 1: Experiment one, concatenation of recorded digits**

The experiment assumed that the impostor could record their voice saying individual numbers and then edit those recordings together to create the correct sequence of numbers for verification / to deceive the SV system.

**Experiment 2: Re-Synthesis**

Re-synthesis techniques are aimed to make the impostor's voice sound closer to the client's voice, assuming that using the client's own voice would be more effective in deceiving the system.

*Re-synthesis 1:* They analysed the sound properties and the pitch of the speech to figure out how to recreate it. They used a special synthesizer that can generate different types of sounds like vowels, nasal sounds, and sounds without voice. They adjusted the synthesizer's settings based on the analysis to make it sound like the original speech.

*Re-synthesis 2:* They extracted important features like pitch, energy, and formant frequencies from the original speech. To make the synthesized speech sound more like the original, they adjusted the formant frequencies by comparing the synthetic and original speech on a frame-by-frame basis.

**Experiment 3: Diphone Synthesis**

A commercial synthesis system - infovox 330 was used to create the desired utterances (speech). Two scenarios were tested: one where isolated digits were concatenated and one where the synthesis system produced the exact requested sequences.

**Results**

The experiment showed that diphone synthesis and concatenation of synthesized digits can be distinguished from the real client's speech. When the synthesized speech was created by joining complete words together, the system was fooled and couldn't differentiate it. The re-synthesis techniques were somewhat better but could still be detected with proper thresholds. The system maintained low error rates even when excluding the synthesized speech attempts.

Using complete words for impostor attempts was highly successful, while the diphone synthesis and re-synthesis techniques were easily detected but didn't cause many false rejects for the real client. Interestingly, choosing other speakers who sounded similar to the client and using their recordings as impostor attempts yielded better results than the re-synthesis techniques tested.

Random testing of speakers accurately predicts SV system performance, but detecting technical impostor attempts is necessary for improved security.

|  |  |  |  |
| --- | --- | --- | --- |
| Detect/Mitigate | Universal | Prior Knowledge | Only Defensive? |
| Evaluates vulnerability of SV system. | Yes, as it can be distinguished. | Impostor has prior knowledge about the system. SV system doesn’t necessarily require prior training. | No, it aims to understand and address the risks associated with intentional impostor attacks. |

|  |  |  |  |
| --- | --- | --- | --- |
| Add Ons | Modified Network | Test on Attack Knowledge | Attack |
| Hidden Markov Model (HMM) based speaker verification system and the parameterization of speech using LPCC coefficients. | - | Intentional impostor has extensive knowledge about the person they are trying to deceive as well as the system they are attacking. | Re-Synthesis, Concatenation of whole words, Diphone Synthesis. |

|  |  |  |
| --- | --- | --- |
| Architecture | Data Set | Accuracy |
| Whole word left-right HMMs with 2 states per phoneme and 2 mixtures per state. LPCC coefficients are used for speech parameterization. | Telephone quality speaker verification database | Choosing other speakers who sounded similar to the client and using their recordings as impostor attempts yielded better results than the re-synthesis techniques tested. |