Tubes Among US: Analog Attack on Automatic Speaker Identification

“Speak, so that I may identify you.”

• Our voices are distinct

• Voice biometrics can identify people

This Technology is called Speaker Identification
Speaker Identification
Applications of Speaker Identification

• Voice-enabled devices; Siri
  • Integrity
  • Personalization

• Phone banking
  • Seamless Identification and Authentication
Security-critical applications

Your voice is your password.
Moving at the speed of life. Yours.

With Voice ID, we can verify you by the sound of your voice.

Similar to a fingerprint, Voice ID uses your unique voiceprint to verify you—so it's easy, fast and secure.
Is voice as secure as a password or a fingerprint?
Securing Speaker Identification against attacks

- Speaker Identification + Liveness detection
Liveness Detection

- **Assumption**: Voice is authentic *if* it comes from a human

---

**How is it safer?**

Fraudsters and hackers may be able to steal or guess your passwords, but they can’t replicate your voice. We measure the mechanics of how sounds are produced, rather than the sounds themselves. **Voice ID is sensitive and sophisticated enough to detect if someone is impersonating you or playing a recording** – and recognize you, even if you have a cold or sore throat.
Breaking the assumption

• What if there is an attack that is *not synthetic*
• An attack that can reshape one’s voice to sound like another

Still a human voice, not synthetic
Mystique

Acoustic Environment

System Under Attack

Human adversary
Tube
Speaker Identification
Liveness Detection
Live voice
digital speaker
Reject access
id
Lea
Why it works?

The human vocal tract is a resonator

A tube is a controllable extension of the vocal tract
Experiment Setup

• Lots of unsuccessful trials!
Experiment Setup: The one that works

➢ Ask me what the tube filter look like in the measurements
Evaluation pipeline

Evaluation at Scale on Celebrities recordings dataset

Live Human Evaluation

Liveness Detection

2 pre-trained speaker identification models
Human Evaluation of Mystique

- 14 participants: 8 male and 6 female
- Using Mystique to impersonate celebrities

<table>
<thead>
<tr>
<th>ID</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
</tr>
</thead>
<tbody>
<tr>
<td>sex</td>
<td>F</td>
<td>M</td>
<td>M</td>
<td>F</td>
<td>F</td>
<td>M</td>
<td>M</td>
<td>F</td>
<td>M</td>
<td>M</td>
<td>F</td>
<td>M</td>
<td>M</td>
<td>F</td>
</tr>
<tr>
<td>Success Rate (%)</td>
<td>56</td>
<td>66</td>
<td>72</td>
<td>74</td>
<td>65</td>
<td>49</td>
<td>60</td>
<td>75</td>
<td>47</td>
<td>53</td>
<td>70</td>
<td>61</td>
<td>35</td>
<td>78</td>
</tr>
</tbody>
</table>

- Attack success rate ranges from 35% to 75%

➢ Ask me about the baseline impersonation success rate (without Mystique)
Liveness detection

• State-of-the-art liveness detection methods fail to detect Mystique as an attack

<table>
<thead>
<tr>
<th>Model</th>
<th>EER</th>
<th>FAR @ FRR=0</th>
</tr>
</thead>
<tbody>
<tr>
<td>LA-LCNN</td>
<td>30%</td>
<td>77%</td>
</tr>
<tr>
<td>PA-LCNN</td>
<td>31%</td>
<td>99%</td>
</tr>
<tr>
<td>Void-SVM</td>
<td>62%</td>
<td>98%</td>
</tr>
<tr>
<td>Void-DNN</td>
<td>35%</td>
<td>92%</td>
</tr>
<tr>
<td>Void-LCNN</td>
<td>33%</td>
<td>93%</td>
</tr>
</tbody>
</table>

Lower is better

• ML models fit to their training data
  • A tube as an analog acoustic filter falls outside their training distribution

➢ Ask me about our attempts to break Mystique
Conclusion

• Speaker identification models make assumptions about the physical world
  • Expected noise and reverberation
  • Hardware

• These assumptions do not always hold in real-life
  • The analog space is very diverse

• Adversaries can exploit these assumptions to implement ML services

• Biometric authentication is not as secure as advertised