Evaluating the Effectiveness of Protection Jamming Devices in Mitigating Smart Speaker Eavesdropping Attacks Using Gaussian White Noise

Payton Walker
Nitesh Saxena

SPIES Lab @ Texas A&M University

{prw0007, nsaxena}@tamu.edu

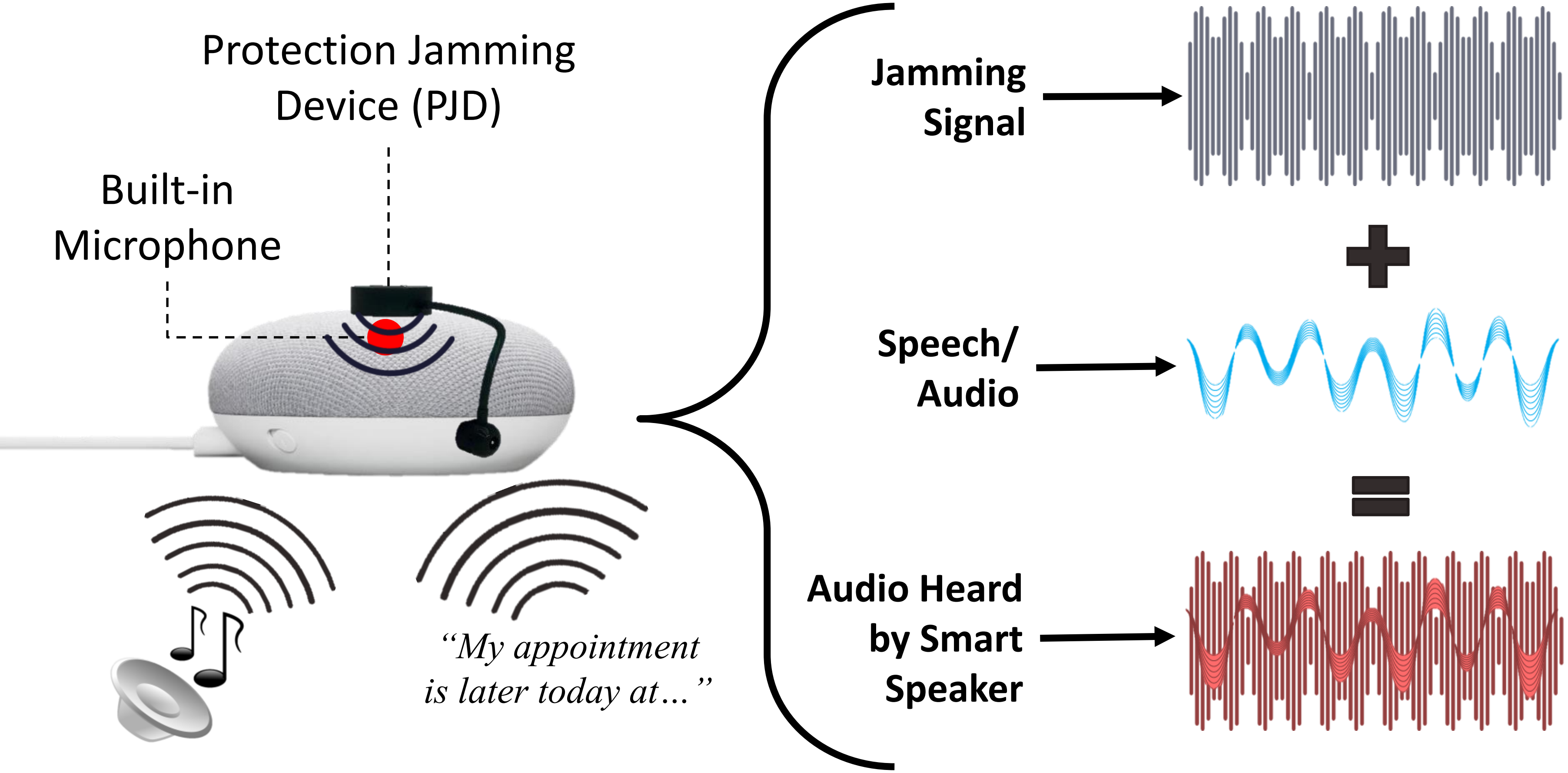
Voice Controllable Systems

amazon alexa

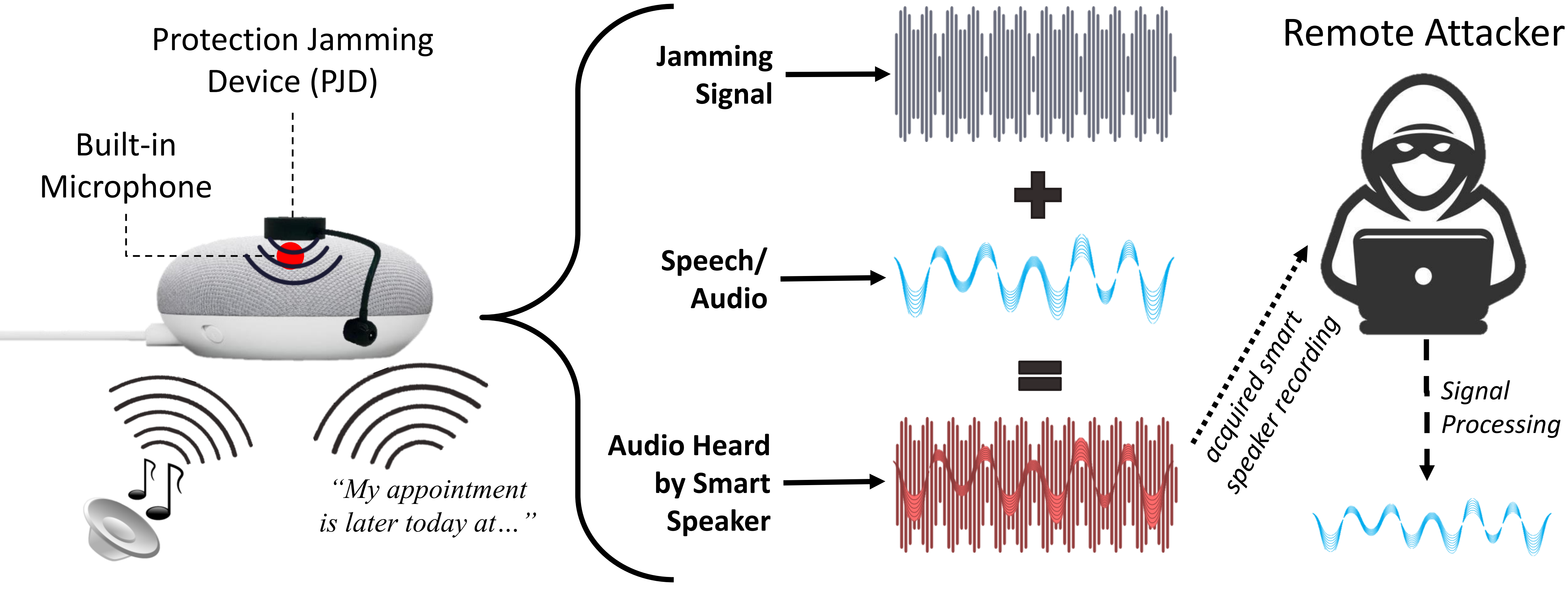
Google Assistant

[Cohen, PCMag ‘20]

Protection Jamming Devices


Protection Jamming Device (PJD)
Built-in Microphone
“My appointment is later today at…”
Jamming Signal
Speech/ Audio
Audio Heard by Smart Speaker

Protection Jamming Devices


Protection Jamming Device (PJD)
Built-in Microphone
"My appointment is later today at…"
Jamming Signal
Speech/ Audio
Audio Heard by Smart Speaker
Remote Attacker
acquired smart speaker recording
Signal Processing

Motivation

1. *Are Protection Jamming Devices always effective at mitigating smart speaker eavesdropping?*

2. *Specifically, how well would Gaussian White Noise perform as the jamming signal?*

Our Contributions

Provide an overview of existing Protection Jamming Device implementations and other related works (*found in paper*).

Built our own Protection Jamming Device modeled after existing implementations and conduct experiments to build a dataset of smart speaker recordings of speech in the presence of a jamming signal.

Signal/speech quality analysis including time and frequency domain inspection, cross-correlation, and use of quality metrics such as Signal-to-Noise Ratio (SNR), and Perceptual Evaluation of Speech Quality (PESQ).

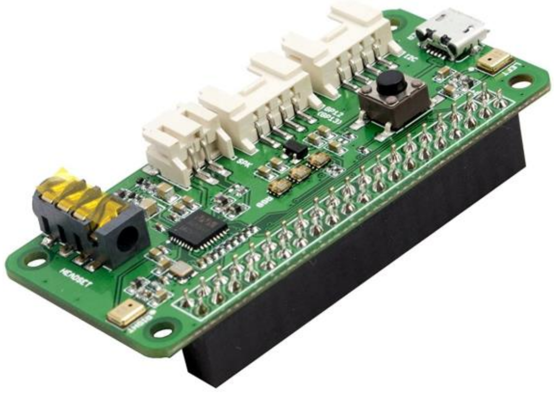
Implement machine learning to demonstrate speech (digit) recognition, speaker and gender identification, and song recognition.

Contents

- Protection Jamming Device Implementation
- Details and Methodology for Data Collection
- Signal Analysis and Observations
- Speech Recognition and Results
- Summary and Conclusions

Implementation of PJD

- Design based on open-source instructions for Project Alias*
  - Raspberry Pi3 + ReSpeaker 2-Mics Pi HAT expansion board
  - JST 2.0 connector + 16mm tiny speaker



- Modifications:
  - Source Code
  - Tiny Speaker Placement

- Determined volume for injected noise and confirmed functionality.

** [Karmann, Online '18]*

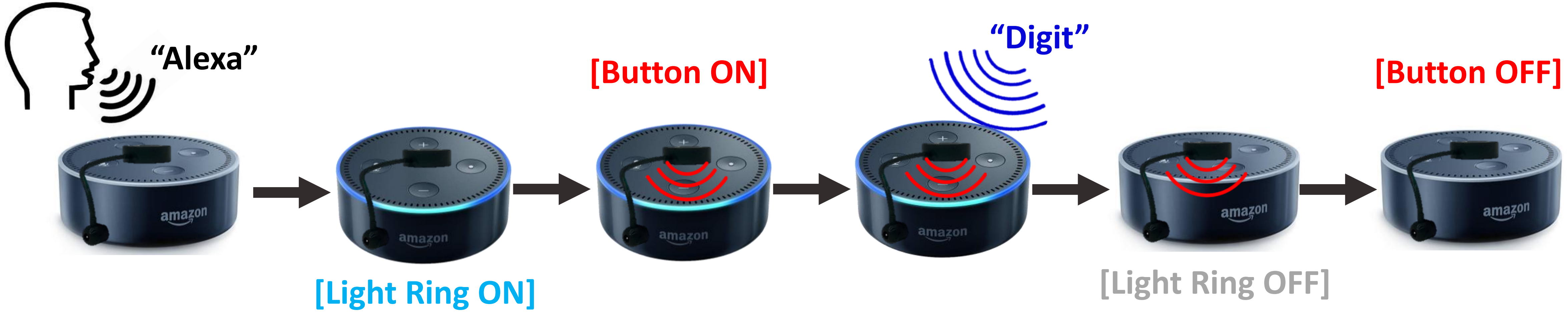
Experimental Details

Smart Speaker Device:	**Amazon Echo Dot**
Speech Source:	**SRS-XB2 Bluetooth Speaker** (0.5 m distance)
Speech Loudness (SPL):	**60, 65, 70 dB** (confirmed with digital sound level meter)
Speech Samples:	**TIDIGITS dataset** (digits 0-9), 5 Female and 5 male speakers
Jamming Signal:	Gaussian white noise (generated using Audacity software)
# Samples Collected:	**5** samples X **10** speaker X **10** digits X **3** SPL levels = **1,500** recordings*

**recordings recovered from Alexa Voice History*

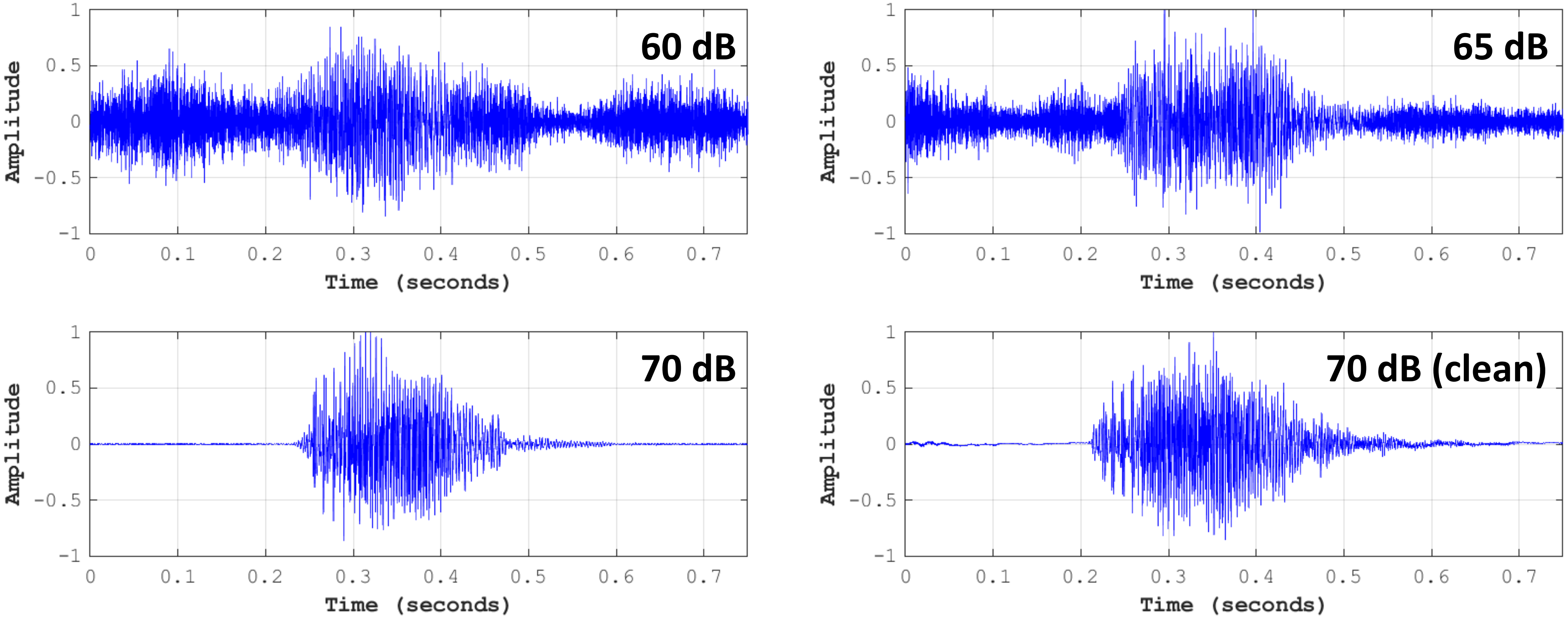
Data Collection Steps

1. Manually activate Echo Dot using the wake word, “Alexa”.
2. After activation (blue light ring), button on PJD is pressed to start the jamming signal.
3. Speech sampled played from Bluetooth speaker.
4. Continue jamming until smart speaker stops recordings (light ring powers down).
5. Stop jamming by pressing the button on the PJD device.
6. Recording is saved (.wav) from Alexa Voice History → post-processing (VOICEBOX)



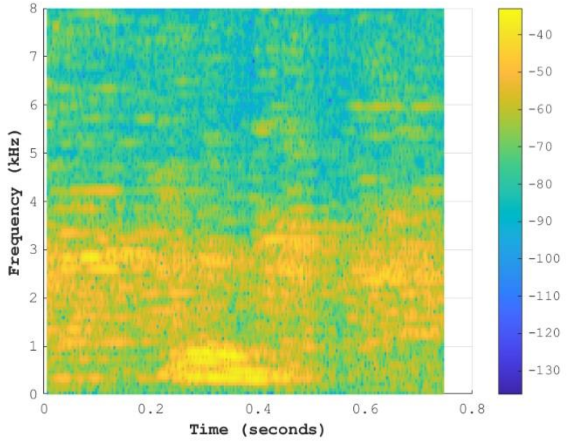
Signal Analysis – Time Domain

{Speaker ID: FAC, Digit: “One”}

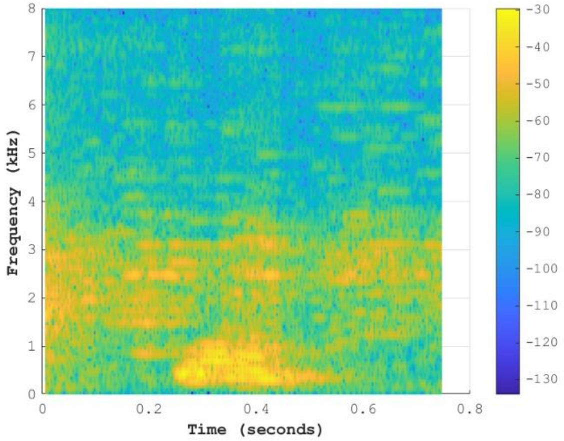


Signal Analysis – Frequency Spectrum

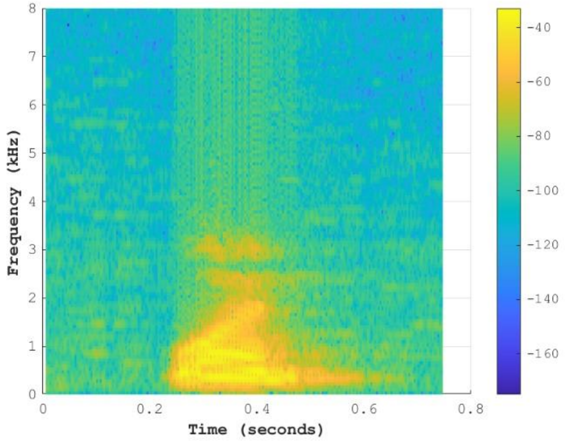
{Speaker ID: FAC, Digit: “One”}



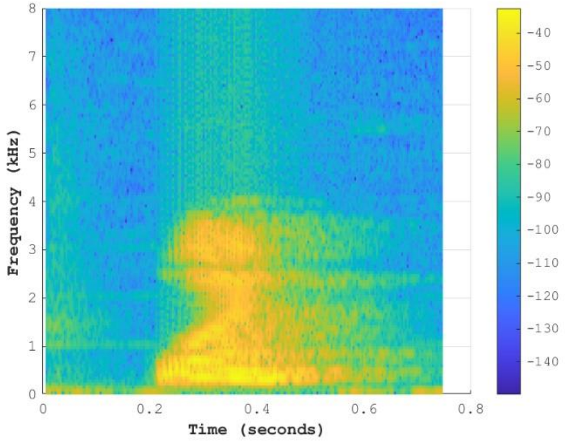
60 dB



65 dB



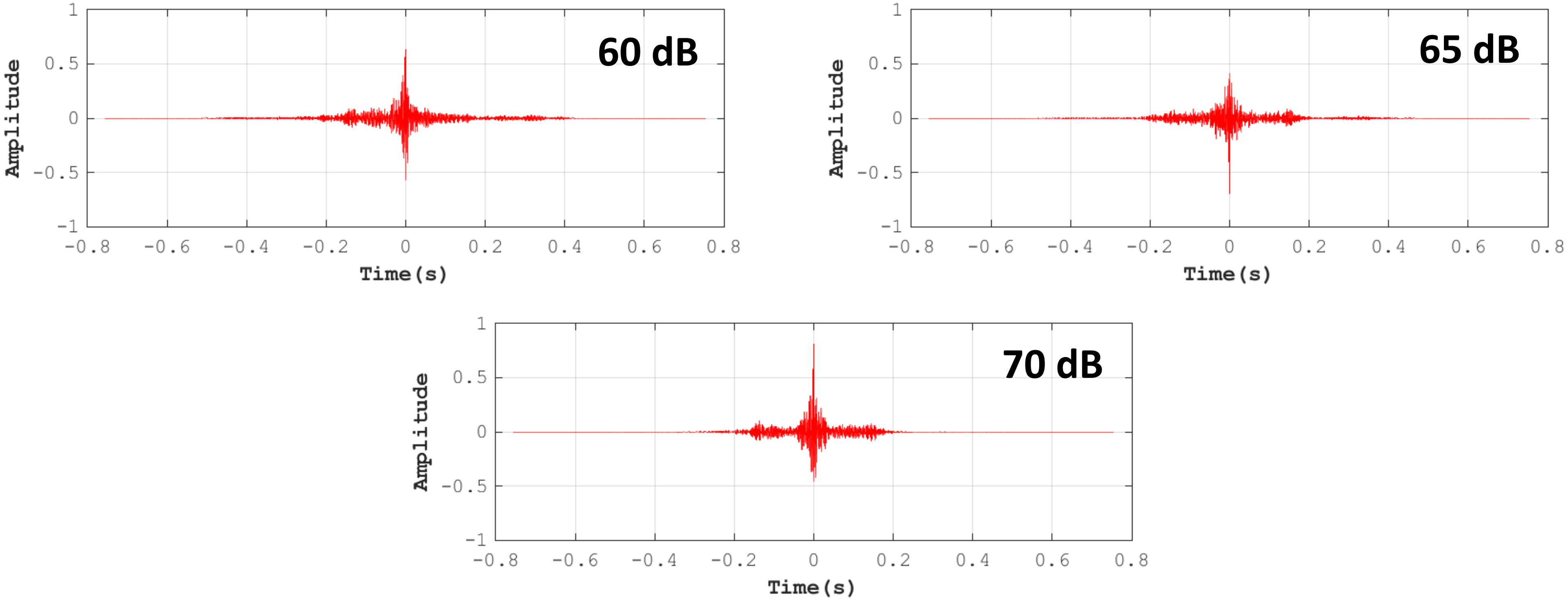
70 dB



70 dB (clean)

Signal Analysis – Cross Correlation

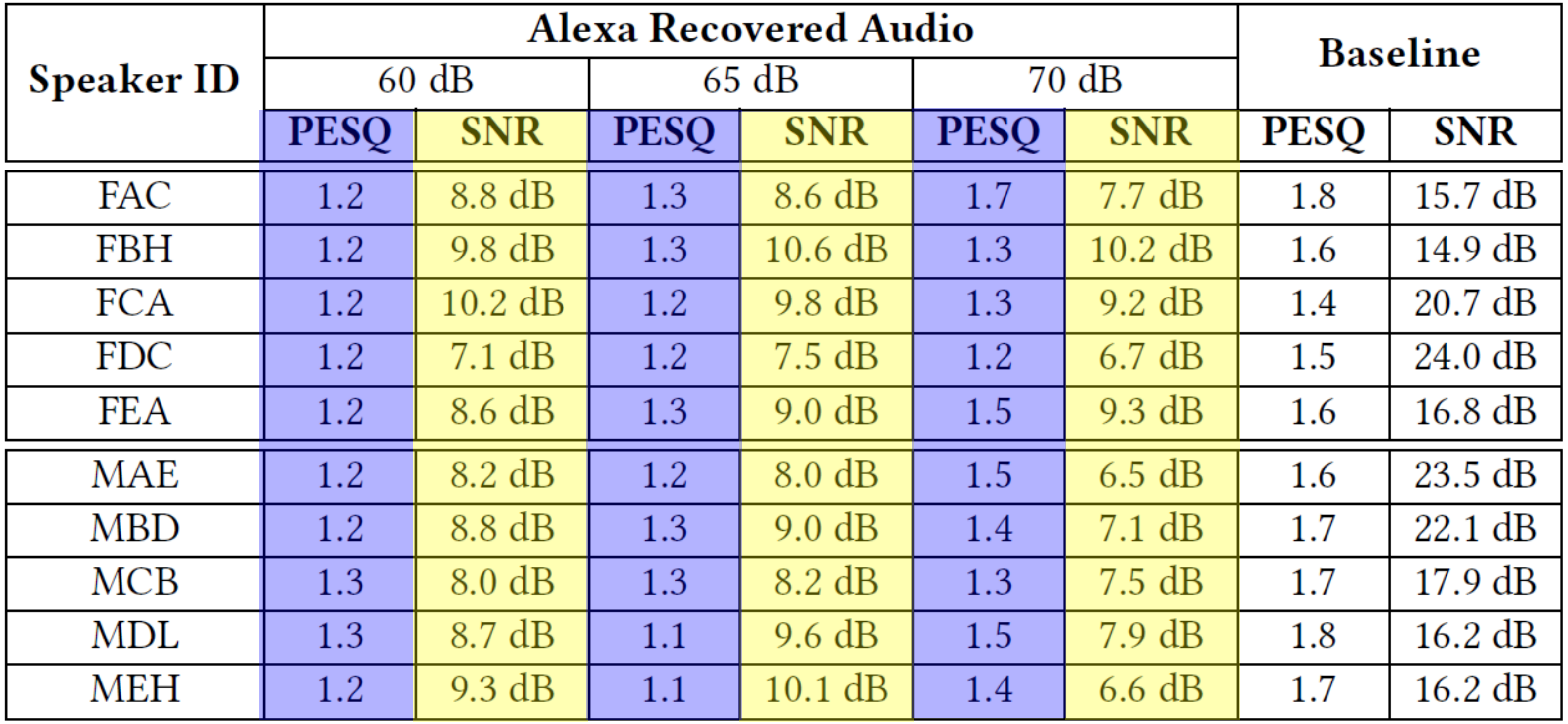
{Speaker ID: FAC, Digit: “One”}



Signal Analysis – Quality Metrics

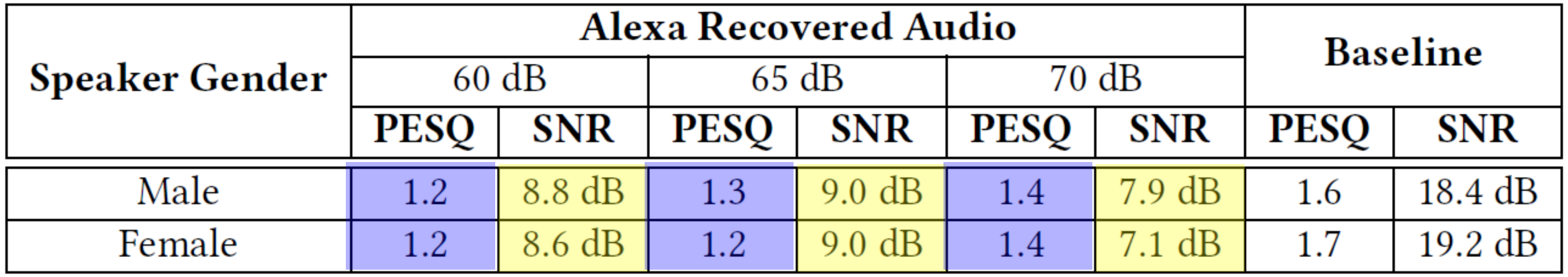
PESQ : Perceptual Evaluation of Speech Quality

SNR: Signal-to-Noise Ratio



Speaker ID	Alexa Recovered Audio						Baseline	
	60 dB		65 dB		70 dB			
	PESQ	SNR	PESQ	SNR	PESQ	SNR	PESQ	SNR
FAC	1.2	8.8 dB	1.3	8.6 dB	1.7	7.7 dB	1.8	15.7 dB
FBH	1.2	9.8 dB	1.3	10.6 dB	1.3	10.2 dB	1.6	14.9 dB
FCA	1.2	10.2 dB	1.2	9.8 dB	1.3	9.2 dB	1.4	20.7 dB
FDC	1.2	7.1 dB	1.2	7.5 dB	1.2	6.7 dB	1.5	24.0 dB
FEA	1.2	8.6 dB	1.3	9.0 dB	1.5	9.3 dB	1.6	16.8 dB
MAE	1.2	8.2 dB	1.2	8.0 dB	1.5	6.5 dB	1.6	23.5 dB
MBD	1.2	8.8 dB	1.3	9.0 dB	1.4	7.1 dB	1.7	22.1 dB
MCB	1.3	8.0 dB	1.3	8.2 dB	1.3	7.5 dB	1.7	17.9 dB
MDL	1.3	8.7 dB	1.1	9.6 dB	1.5	7.9 dB	1.8	16.2 dB
MEH	1.2	9.3 dB	1.1	10.1 dB	1.4	6.6 dB	1.7	16.2 dB

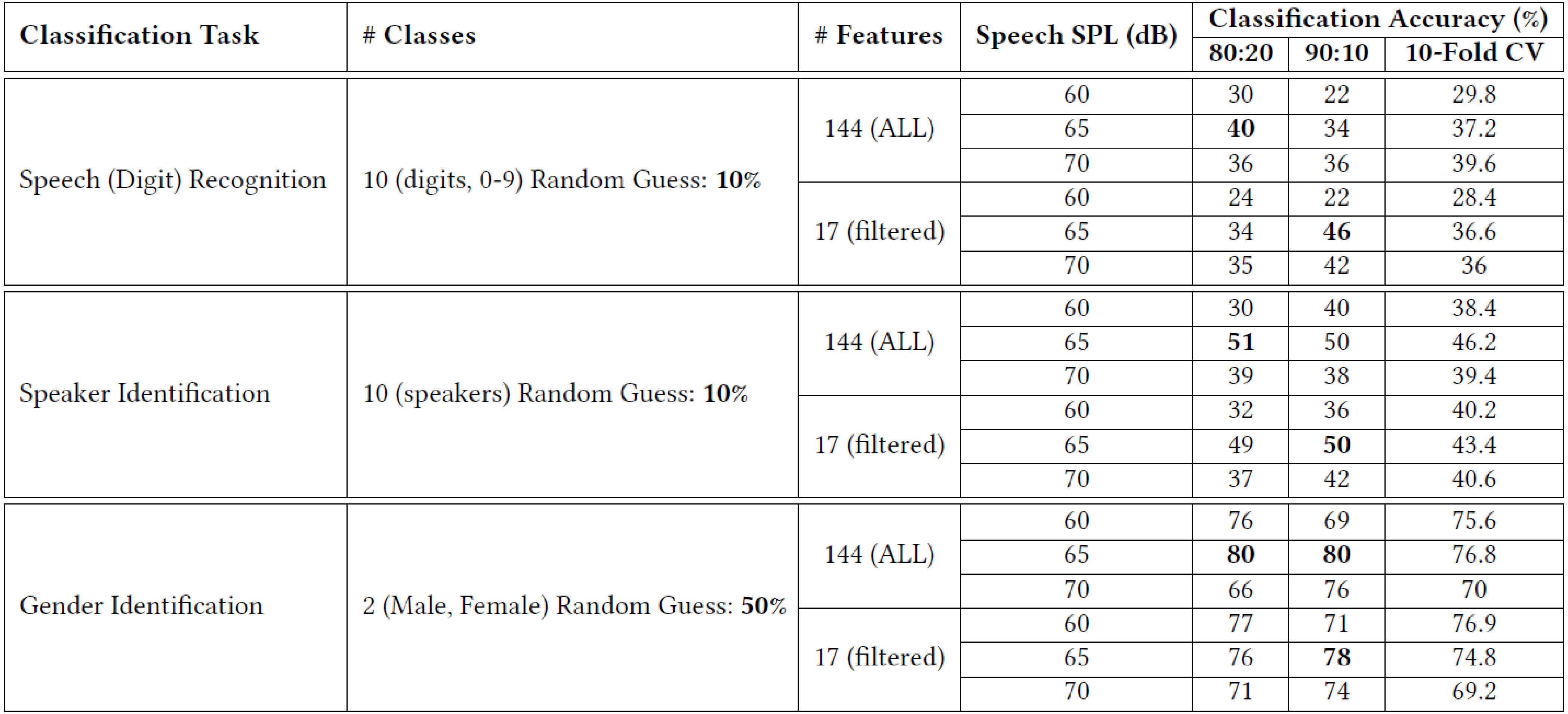
Averaged results from PESQ and SNR analysis for each individual speaker.



Speaker Gender	Alexa Recovered Audio						Baseline	
	60 dB		65 dB		70 dB			
	PESQ	SNR	PESQ	SNR	PESQ	SNR	PESQ	SNR
Male	1.2	8.8 dB	1.3	9.0 dB	1.4	7.9 dB	1.6	18.4 dB
Female	1.2	8.6 dB	1.2	9.0 dB	1.4	7.1 dB	1.7	19.2 dB

Averaged results from PESQ and SNR analysis for both speaker genders.

Speech/Speaker Classification Results



Classification Task	# Classes	# Features	Speech SPL (dB)	Classification Accuracy (%)		
				80:20	90:10	10-Fold CV
Speech (Digit) Recognition	10 (digits, 0-9) Random Guess: **10%**	144 (ALL)	60	30	22	29.8
			65	**40**	34	37.2
			70	36	36	39.6
		17 (filtered)	60	24	22	28.4
			65	34	**46**	36.6
			70	35	42	36
Speaker Identification	10 (speakers) Random Guess: **10%**	144 (ALL)	60	30	40	38.4
			65	**51**	50	46.2
			70	39	38	39.4
		17 (filtered)	60	32	36	40.2
			65	49	**50**	43.4
			70	37	42	40.6
Gender Identification	2 (Male, Female) Random Guess: **50%**	144 (ALL)	60	76	69	75.6
			65	**80**	**80**	76.8
			70	66	76	70
		17 (filtered)	60	77	71	76.9
			65	76	**78**	74.8
			70	71	74	69.2

Classification results (using RandomForest) for Speech (Digit) Recognition and Speaker/Gender Identification.

Song Recognition Results

- Samples from beginning, middle, and end of two songs were isolated (5 seconds each).
  - “Smooth” by Santana and “Blinding Lights” by The Weeknd
- Songs were played at 70 dB and processed in the same way as the speech samples.
- Shazam App was used to attempt song recognition.
  - efficient, scalable, noise and distortion resistant song identification algorithm
- **Recognition Accuracy = 100%** was achieved.

SHAZAM®

Summary and Conclusion

- Custom implementation of a Protection Jamming Device.
- Performed Signal Analysis (time/frequency domain, cross correlation, quality metrics).
- Demonstrate successful Speech Classification tasks.
  - Speech Recog. (**46%**), Speaker Ident. (**51%**), Gender Ident. (**80%**), Song Recog. (**100%**).

➢ ***We conclude that Gaussian white noise is not an effective jamming signal to use in a Protection Jamming Device because speech and speaker information can still be compromised by an attacker.***

- ***Future work to develop jamming signals should focus on more complex noise types because standard noises cannot guarantee privacy of user speech.***