TRANSMISSION LINE COCHLEAR MODEL BASED AM-FM FEATURES FOR REPLAY ATTACK DETECTION

Tharshini Gunendradasan[1], Saad Irtza[1], Eliathamby Ambikairajah[1,2], Julien Epps[1,2]

[1] The School of Electrical Engineering and Telecommunications, The University of New South Wales, Sydney, Australia

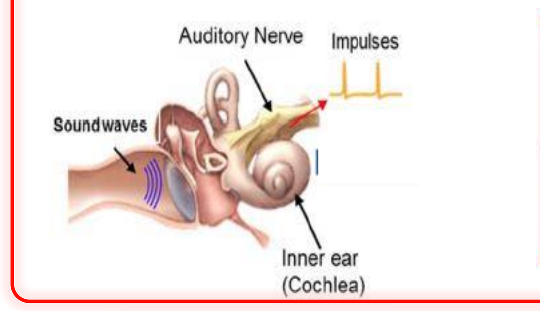
[2] Data61, CSIRO, Australia

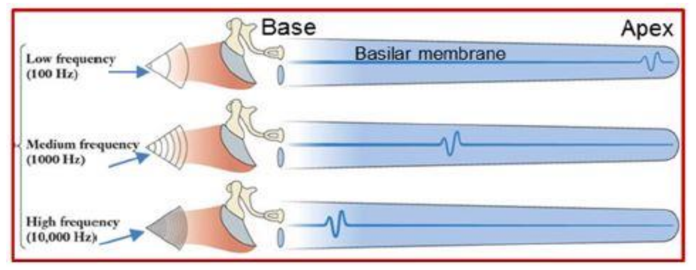
tharshini.gunendradasan@student.unsw.edu.au, s.irtza@unsw.edu.au, e.ambikairajah@unsw.edu.au, j.epps@unsw.edu.au

1. Motivation

- Humans are excellent spectrum analyzers and it is hypothesized that some properties of the human auditory system, specifically high roll-off magnitude response, are beneficial in identifying spoofing attacks
- A Transmission Line Cochlear (TLC) model which more accurately resembles the human auditory system is proposed for the front-end of feature extraction for replay spoofing attack detection
- Amplitude Modulation (AM) and Frequency Modulation (FM) features are extracted from the TLC front-end

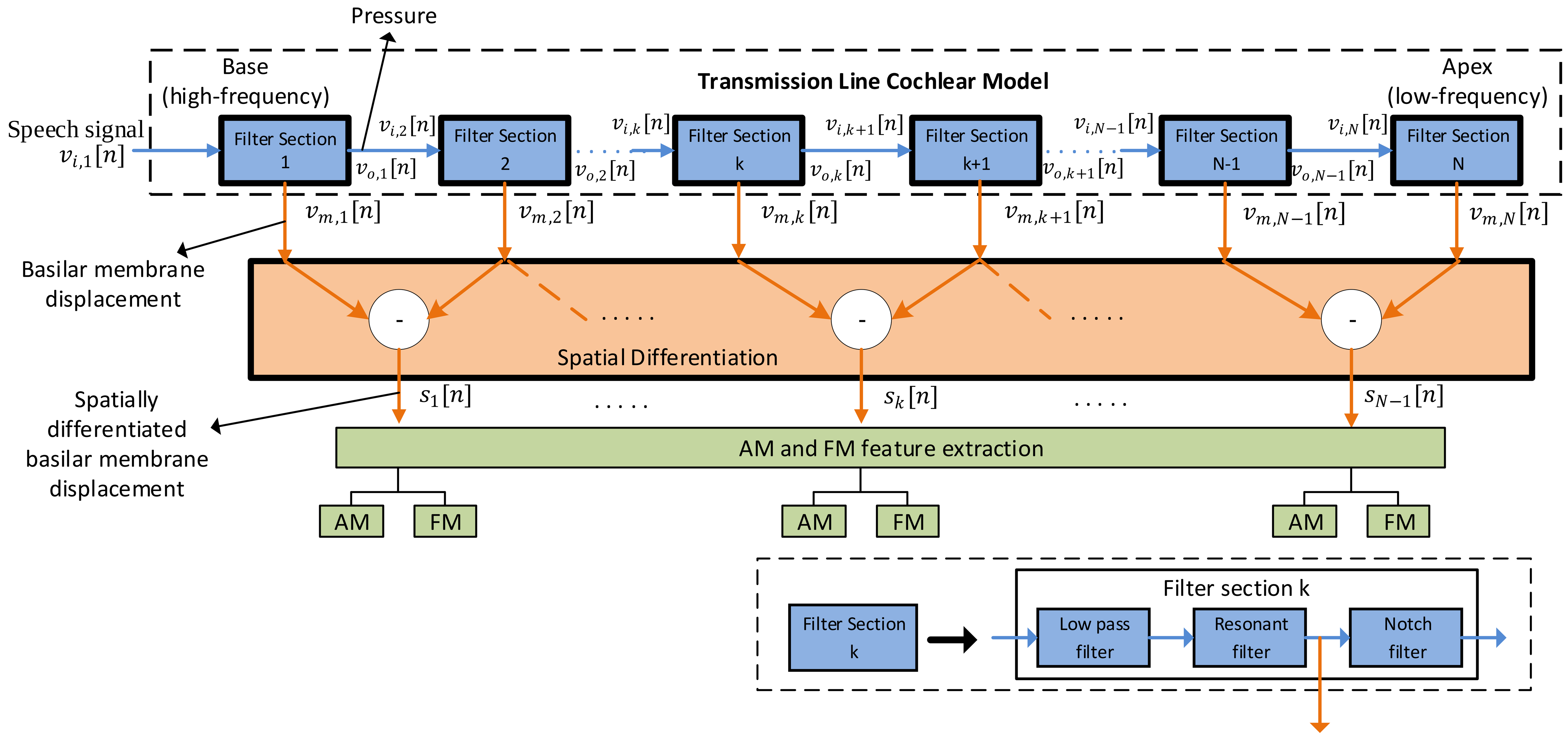
2. Human auditory system

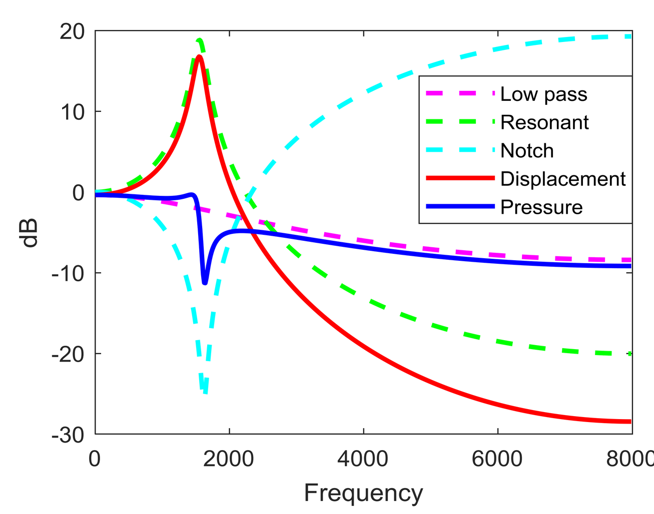




- The **inner ear** consists of the cochlea, responsible for converting the vibrations of sound waves into electrochemical impulses
- The basilar membrane resonates according to the frequency content of the incoming speech signal. For low frequency the apex will resonate, the base will resonate for high frequency

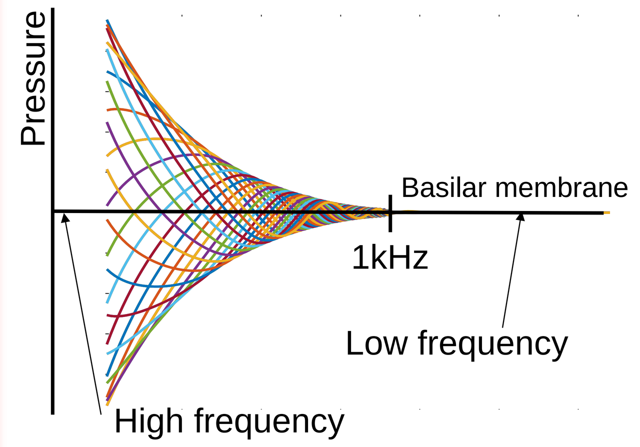
3. Transmission line cochlear model



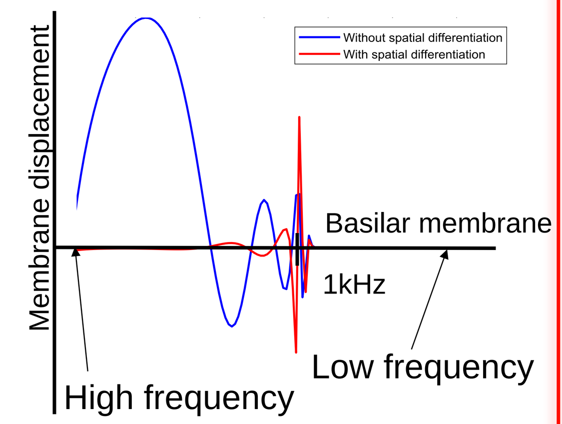


- Pressure along the basilar membrane is represented using a set of filters cascaded in series
- Each filter section consists of lowpass, resonant and notch filters
  - The pressure transfer function of the filter section k: $H_k(z) = \frac{V_{o,k}(z)}{V_{i,k}(z)} = K \frac{1-a_0}{1-a_0 z^{-1}} \frac{1-b_1+b_2}{1-b_1 z^{-1}+b_2 z^{-2}} \frac{1-a_1 z^{-1}+a_2 z^{-2}}{1-a_1+a_2}$
- Displacement of the basilar membrane is tapped at the intermediate point of the filter section
  - Displacement transfer function: $\frac{V_{k,m}(z)}{V_{k,i}(z)} = K \frac{1-a_0}{1-a_0 z^{-1}} \frac{1-b_1+b_2}{1-b_1 z^{-1}+b_2 z^{-2}}$
- Spatial differentiation is performed on the basilar membrane displacement to represents the fluid coupling effect between the adjacent sections of the membrane
  - Spatial differentiation: $s_k[n] = v_{m,k+1}[n] - v_{m,k}[n]$

4. Pressure and displacement along the basilar membrane for input sinusoidal signal of frequency 1kHz

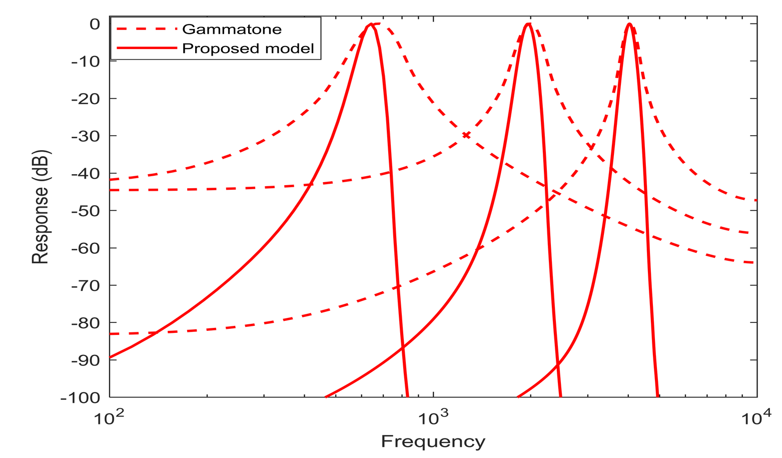


- Pressure along the membrane dies off after 1kHz resonant point

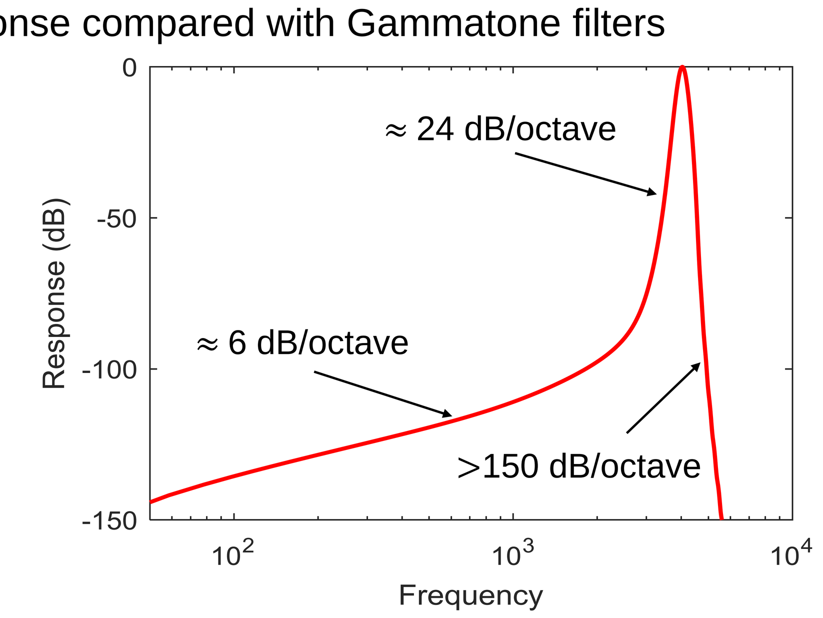


- Basilar membrane resonates with high amplitude at 1kHz resonant point after spatial differentiation

5. Frequency response of transmission line cochlear model and parallel filter bank model

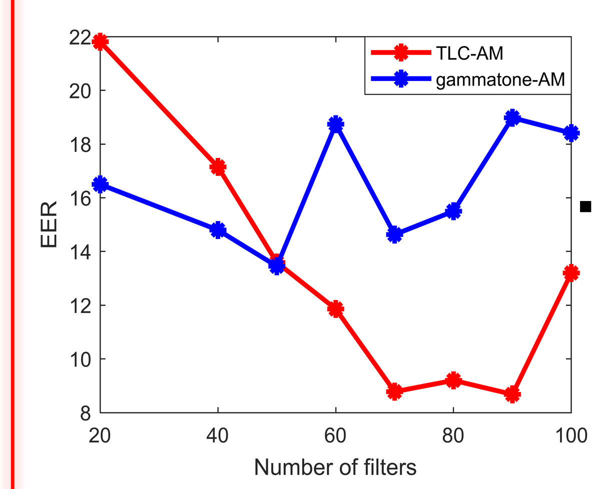


- Transmission line cochlear model have very sharp-off frequency response compared with Gammatone filters



- Frequency response of the proposed cochlear model approximately matches human cochlear measurements

6. Effect of high frequency roll-off of TLC model in spoofing detection



- TLC model gives minimum EER with a large number of filters (N=90), for Gammatone filters this is at a low number of filters (N=50)
- The sharp roll-off achieved by the TLC model helps to capture non-overlapping information from the adjacent bands, which allows utilizing a large number of filters to obtain accurate details in narrow frequency bands

7. Results on ASV spoof 2017 version 1 and version 2 database

		Feature set	EER
AM	Paralel	HT-IACC [22]) (BW), (V1)	19.27
		HT-IACC [22] (GA), (V1)	12.12
		ESA-IACC [22] (BW), (V1)	21.43
		ESA-IACC [22] (GA), (V1)	12.00
		VESA-IACC [19] (GA), (V1)	11.94
		AM-ConvRBM-CC [18] (CV), (V1)	12.76
	TLC	**TLC-AM (V1)**	**8.51**
		TLC-AM (V2)	**8.68**
FM	Paralel	HT-IFCC [22] (BW), (V1)	39.40
		HT-IFCC [22] (GA), (V1)	14.62
		ESA-IFCC [22] (BW), (V1)	28.69
		ESA-IFCC [22] (GA), (V1)	12.79
		VESA-IFCC [19] (BW), (V1)	11.79
		FM- ConvRBM-CC [18] (CV), (V1)	14.96
	TLC	**TLC-FM (V1)**	**10.11**
	TLC	**TLC-FM (V2)**	**11.30**

Fused features set	EER
VESA-IACC+ VESA-IFCC [19], (V1)	7.11
AM-ConvRBM-CC + FM-ConvRBM-CC [18], (V1)	8.89
TLC-AM+ TLC-FM (V1)	**7.32**
TLC-AM+ TLC-FM (V2)	**7.59**
HT-IACC+ HT-IFCC [22], (V1)	10.03
ESA-IACC + ESA-IFCC [22], (V1)	9.64
CQCC [28], (V2)	12.24

8. Conclusion

- The proposed transmission line cochlear model represents the cochlea as a cascade of digital filters
- The TLC model gives steeper high-frequency roll-off compared with parallel filter bank auditory models, capturing information within narrow frequency bands
- Individually both TLC-AM and TLC-FM features show significantly improved performance over other individual AM and FM features extracted from parallel filter banks
- Fusion of TLC-AM and TLC-FM showed improved performance in distinguishing replay attack from genuine speech