https://us05web.zoom.us/j/8967653
7292?pwd=L3NqT1FjMFlQQVRne
G0rNC90c0xqdz09

Exploring Methods of Signal Processing for Keywords Spotting in Hearing Devices

Cypress Payne[1], Jhyv Philor[2], Tadesse Ghirmai[3], Kaibao Nie[3]
[1] Seattle Pacific University, [2] University of Florida, [3] University of Washington Bothell

INTRODUCTION

Problem: Changing settings in hearing aids and cochlear implants can be inconvenient by hand. Using keywords spotting (KWS) allows parameter adjustment to be done by voice.

Current feature extraction methods for KWS:

- Bark spectrum, Mel spectrum, Mel frequency cepstral coefficients (MFCC)

Current deep learning techniques for KWS:

- Recurrent Neural Network (RNN), Convolutional Neural Networks (CNN) , Long-short term Memory (LSTM) [1]

Proposed Method: Applying amplitude modulation on an audio signal to extract information to train a CNN for keywords spotting. Amplitude modulation is vital for speech recognition because at certain frequencies useful speech information can be extracted [2] [3].

We compared the accuracy of a CNN using amplitude modulated signals with different bands versus Mel spectrum-based signal processing.

METHODS

Google speech dataset

- 65000 different utterances of words were split into training, validation, and testing sets [4].
- 11 keywords, consisting of “up”, “down”, and “one” through “nine” were selected for volume control

Preprocessing

- An audio feature extractor in MATLAB was used to convert each audio signal into a Mel spectrogram.
- Amplitude modulation was obtained by passing the signal through multiple contiguous bandpass filters and performing a Hilbert Transform to extract temporal envelopes (Figure 1).

Neural Network

- A CNN with 3 hidden layers was setup in MATLAB (Figure 1).

EXTRACTION

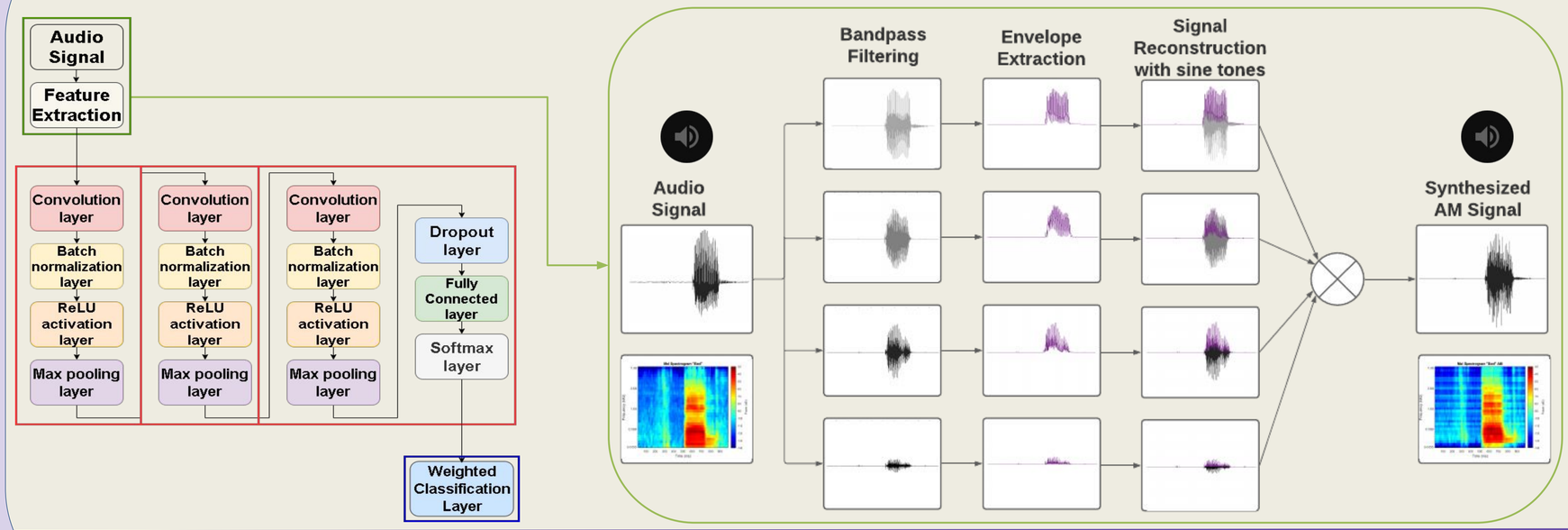


Figure 1: The left of the diagram outlines the entire convolutional neural network, with input, hidden layers, and output. The right displays the feature extraction algorithm that performs amplitude modulation on the signal before it is fed to the neural network.

RESULTS

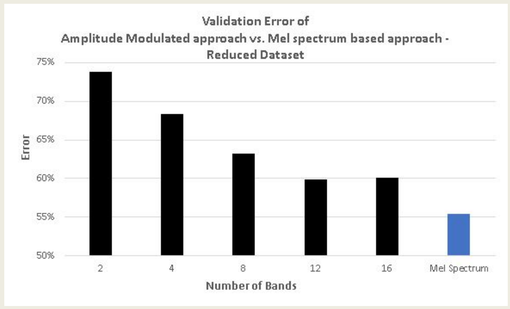


Fig. 2: Accuracy of AM signals with different numbers of bands compared to Mel spectrum accuracy using a reduced dataset

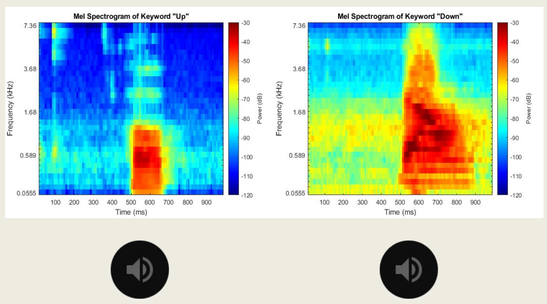


Fig. 3: The left is a Mel spectrogram and recording of the word “up” and the right is a Mel spectrogram and recording of the word “down”.

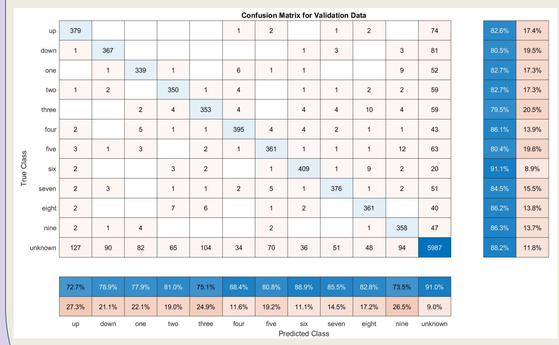


Fig. 4: Confusion matrix of a CNN trained with the full dataset of Mel spectrograms identifying keywords from the validation set
Validation Error: 13.6105%

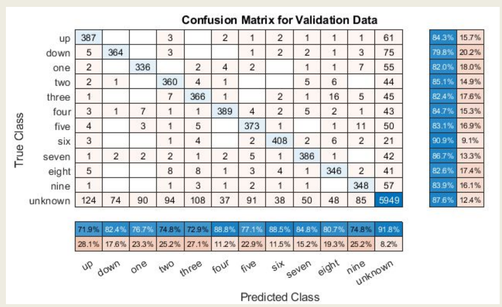


Fig. 5: Confusion matrix of a CNN trained with the full dataset of Mel spectrograms of both original and amplitude modulated signals identifying keywords from the validation set
Validation Error: 13.8085%

CONCLUSIONS

- It is feasible to perform keywords spotting using a convolutional neural network for volume adjustments of hearing devices.
- Mel-based spectral information is vital to train a neural network for keywords spotting, but combining this with amplitude information is a viable method
- Results are consistent with previous studies which show that with 8 or more bands, amplitude modulated audio can be recognized by human subjects.
- More research into amplitude modulation is necessary to determine if solely amplitude modulated signal processing is feasible for keywords spotting systems

REFERENCES

[1] Y. Zhang, N. Suda, L. Lai, and V. Chandra, “Hello Edge: Keyword Spotting on Microcontrollers,” *ArXiv*, vol. abs/1711.07128, 2017.

[2] J. H. Won, C. Lorenzi, K. Nie, X. Li, E. M. Jameyson, W. R. Drennan, and J. T. Rubinstein, “The ability of cochlear implant users to use temporal envelope cues recovered from speech frequency modulation,” *The Journal of the Acoustical Society of America*, vol. 132, no. 2, pp. 1113–1119, 2012.

[3] E. Scheirer and M. Slaney, "Construction and evaluation of a robust multifeature speech/music discriminator," 1997 IEEE International Conference on Acoustics, Speech, and Signal Processing, 1997, pp. 1331-1334.

[4] P. Warden, “Launching the Speech Commands Dataset,” Google AI Blog, 24-Aug-2017. [Online]. Available: https://ai.googleblog.com/2017/08/launching-speech-commands-dataset.html.

ACKNOWLEDGEMENTS

This work was supported by the National Science Foundation REU Award #1757395 at the University of Washington Bothell.

INTRODUCTION

Problem: Changing settings in hearing aids and cochlear implants can be inconvenient by hand. Using keywords spotting (KWS) allows parameter adjustment to be done by voice.

Current feature extraction methods for KWS:

- Bark spectrum, Mel spectrum, Mel frequency cepstral coefficients (MFCC)

Current deep learning techniques for KWS:

- Recurrent Neural Network (RNN), Convolutional Neural Networks (CNN) , Long-short term Memory (LSTM) [1]

Proposed Method: Applying amplitude modulation on an audio signal to extract information to train a CNN for keywords spotting. Amplitude modulation is vital for speech recognition because at certain frequencies useful speech information can be extracted [2] [3].

We compared the accuracy of a CNN using amplitude modulated signals with different bands versus Mel spectrum-based signal processing.

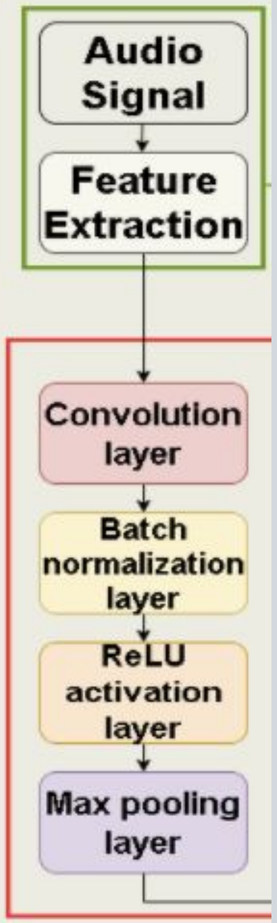


Figure
that pe

We compared the accuracy of a CNN using amplitude modulated signals with different bands versus Mel spectrum-based signal processing.

METHODS

Google speech dataset

- 65000 different utterances of words were split into training, validation, and testing sets [4].
- 11 keywords, consisting of "up", "down", and "one" through "nine" were selected for volume control

Preprocessing

- An audio feature extractor in MATLAB was used to convert each audio signal into a Mel spectrogram.
- Amplitude modulation was obtained by passing the signal through multiple contiguous bandpass filters and performing a Hilbert Transform to extract temporal envelopes (Figure 1).

Neural Network

- A CNN with 3 hidden layers was setup in MATLAB (Figure 1).

RESULTS

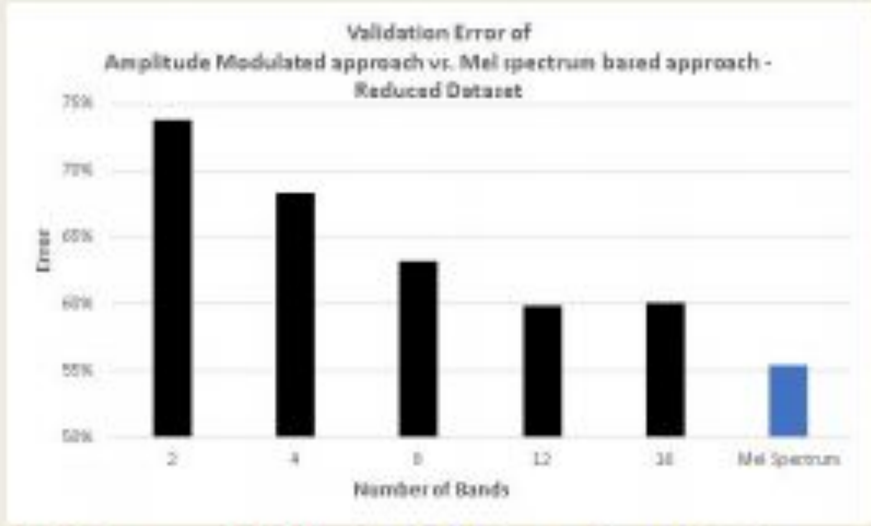


Fig. 2: Accuracy of AM signals with different numbers of bands compared to Mel spectrum accuracy using a reduced dataset

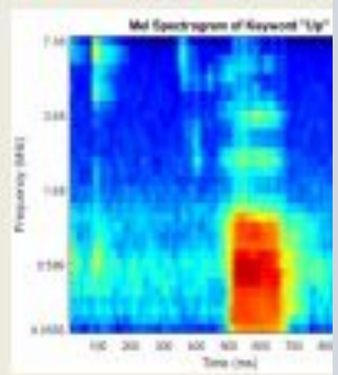


Fig. 3: The left is a Mel sp and the right is a Mel spe

Fig. 4: Confusion matrix of a CNN trained with the full dataset of Mel spectrograms identifying keywords from the validation set Validation Error: 13.6105%

Fig. 5: Confusion matrix spectrograms of both orig identifying keywords fro Validation Error: 13.8085

EXTRACTION

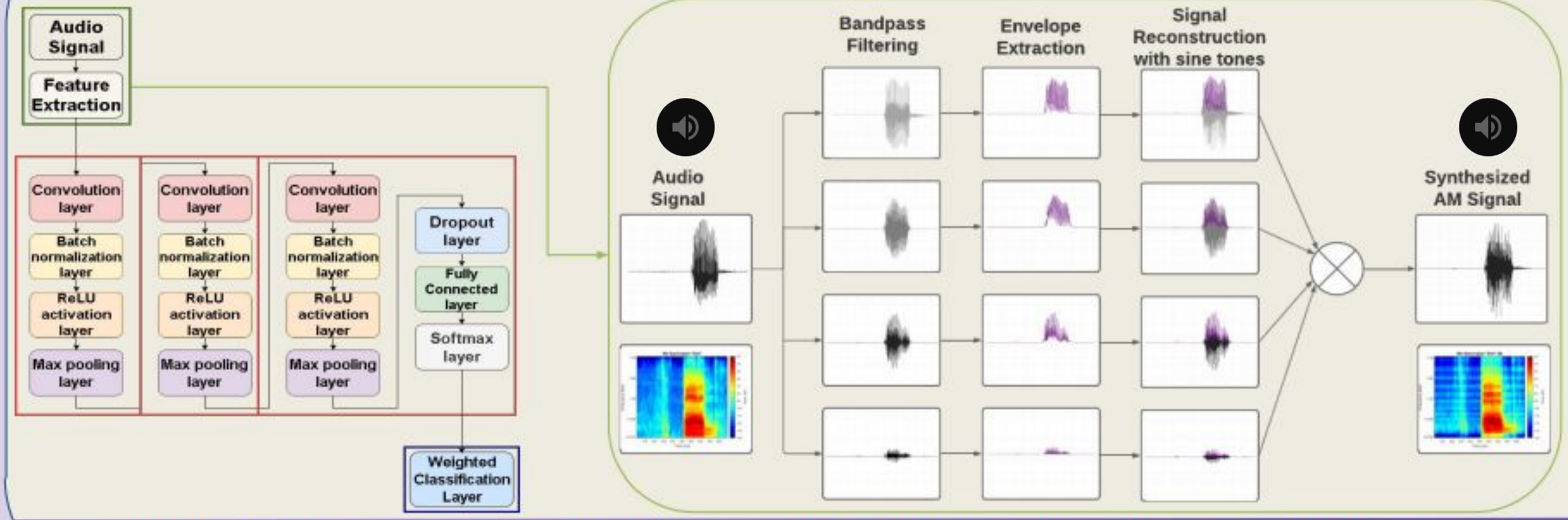
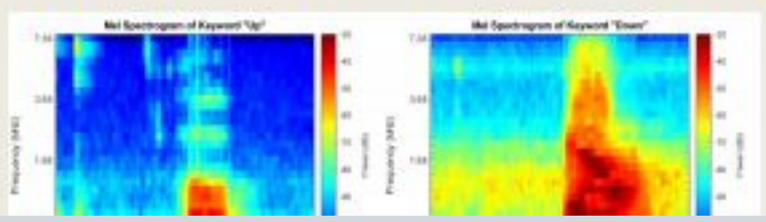


Figure 1: The left of the diagram outlines the entire convolutional neural network, with input, hidden layers, and output. The right displays the feature extraction algorithm that performs amplitude modulation on the signal before it is fed to the neural network.

RESULTS



CONCLUSIONS

- It is feasible to perform keywords spotting using a convolutional neural network for volume adjustments of hearing devices.
- Mel-based spectral information is vital to train a neural network for keywords spotting, but combining this with amplitude information is a viable method
- Results show that with 8 or more bands, amplitude modulated

RESULTS

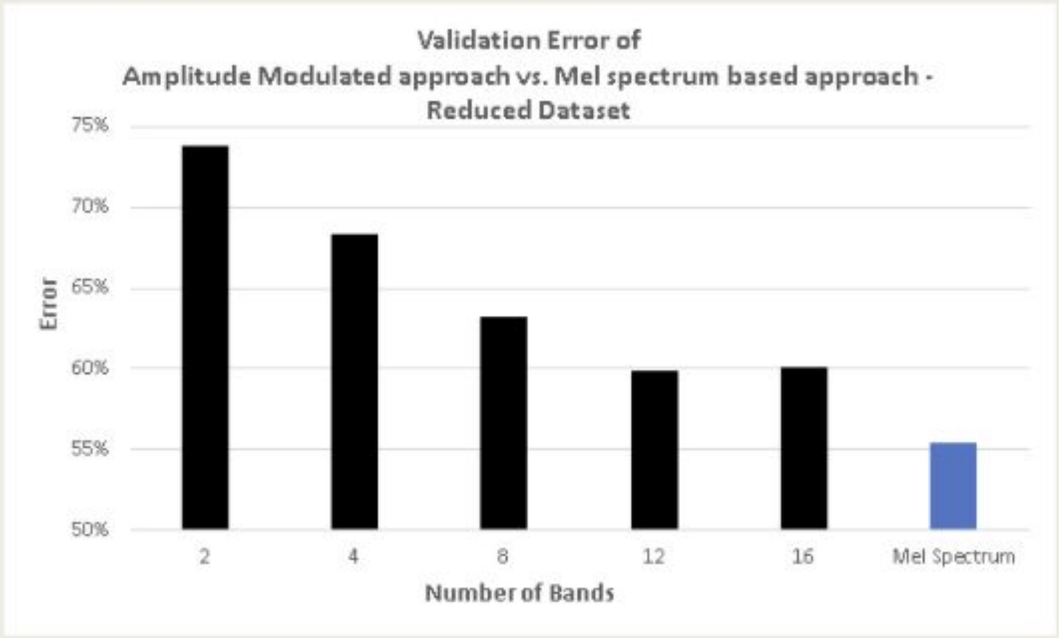


Fig. 2: Accuracy of AM signals with different numbers of bands compared to Mel spectrum accuracy using a reduced dataset

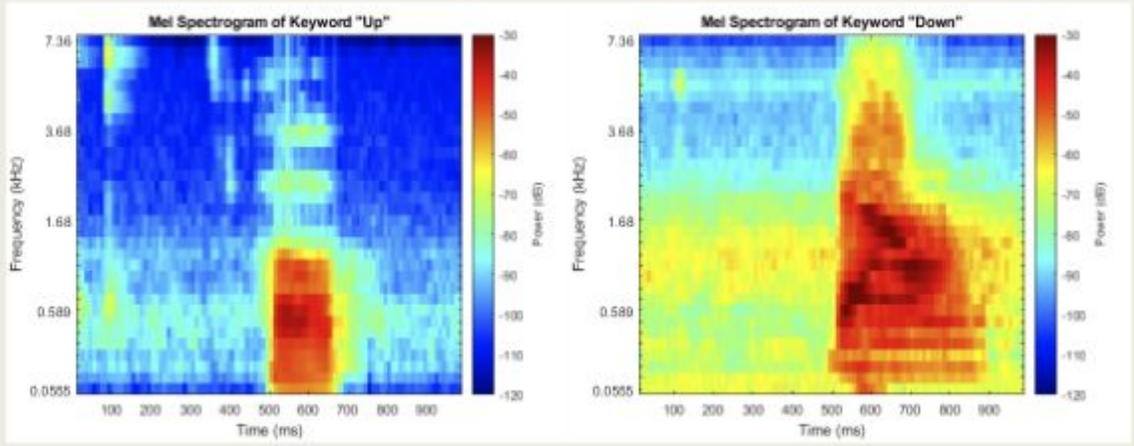


Fig. 3: The left is a Mel spectrogram and recording of the word "up" and the right is a Mel spectrogram and recording of the word "down".

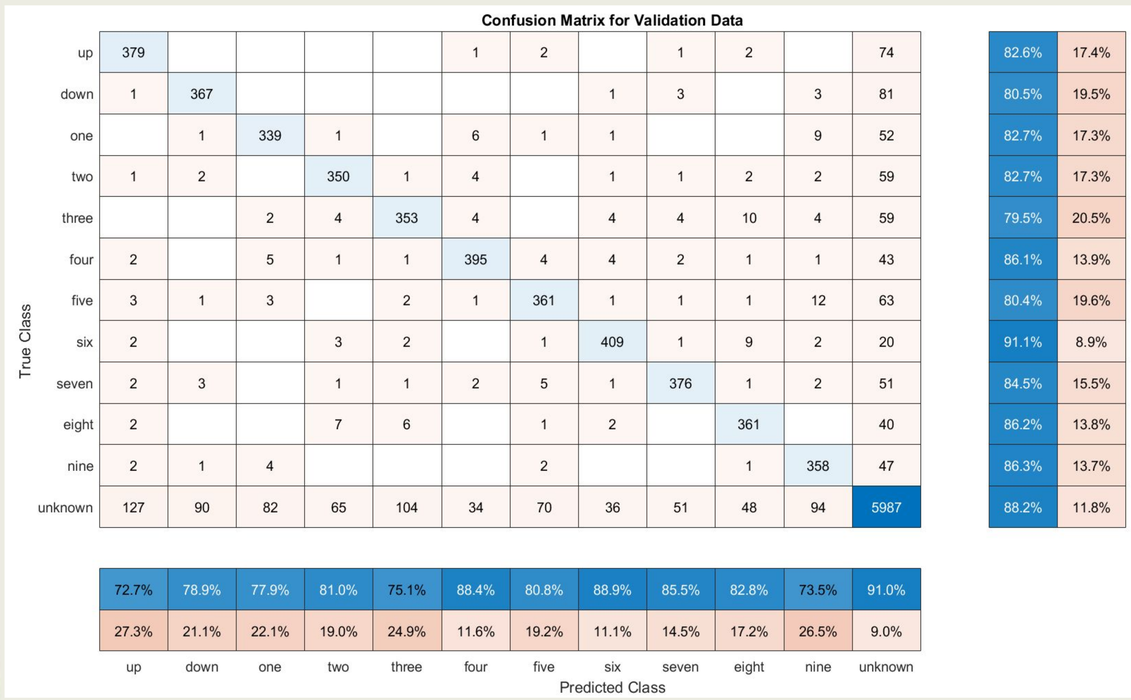


Fig. 4: Confusion matrix of a CNN trained with the full dataset of Mel spectrograms identifying keywords from the validation set

Validation Error: 13.6105%

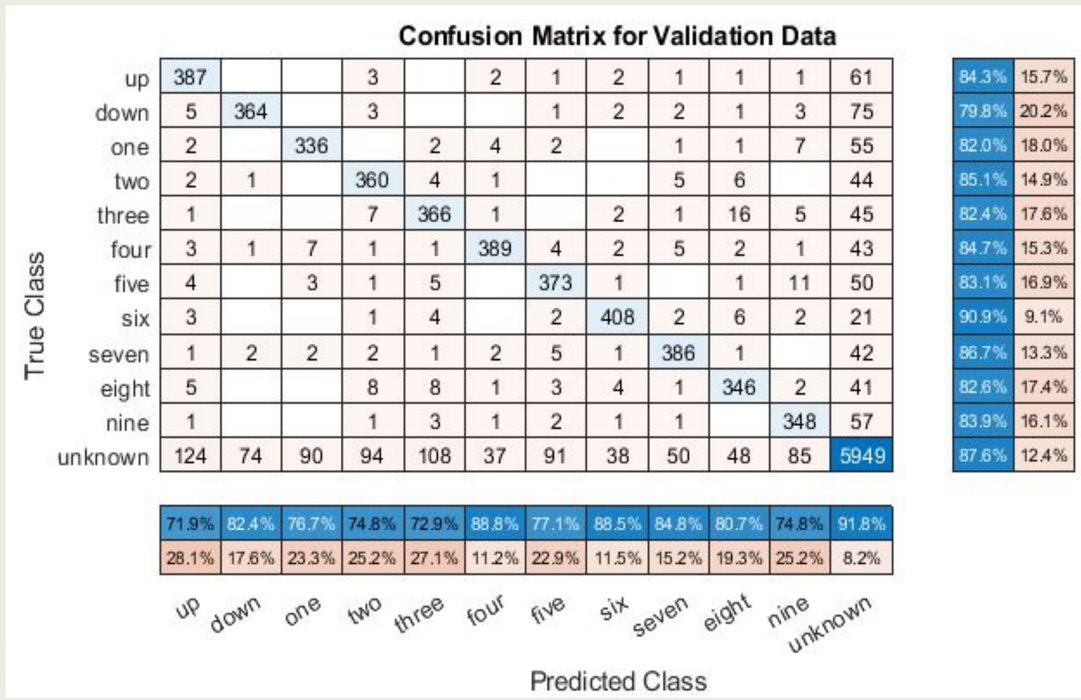


Fig. 5: Confusion matrix of a CNN trained with the full dataset of Mel spectrograms of both original and amplitude modulated signals identifying keywords from the validation set

Validation Error: 13.8085%

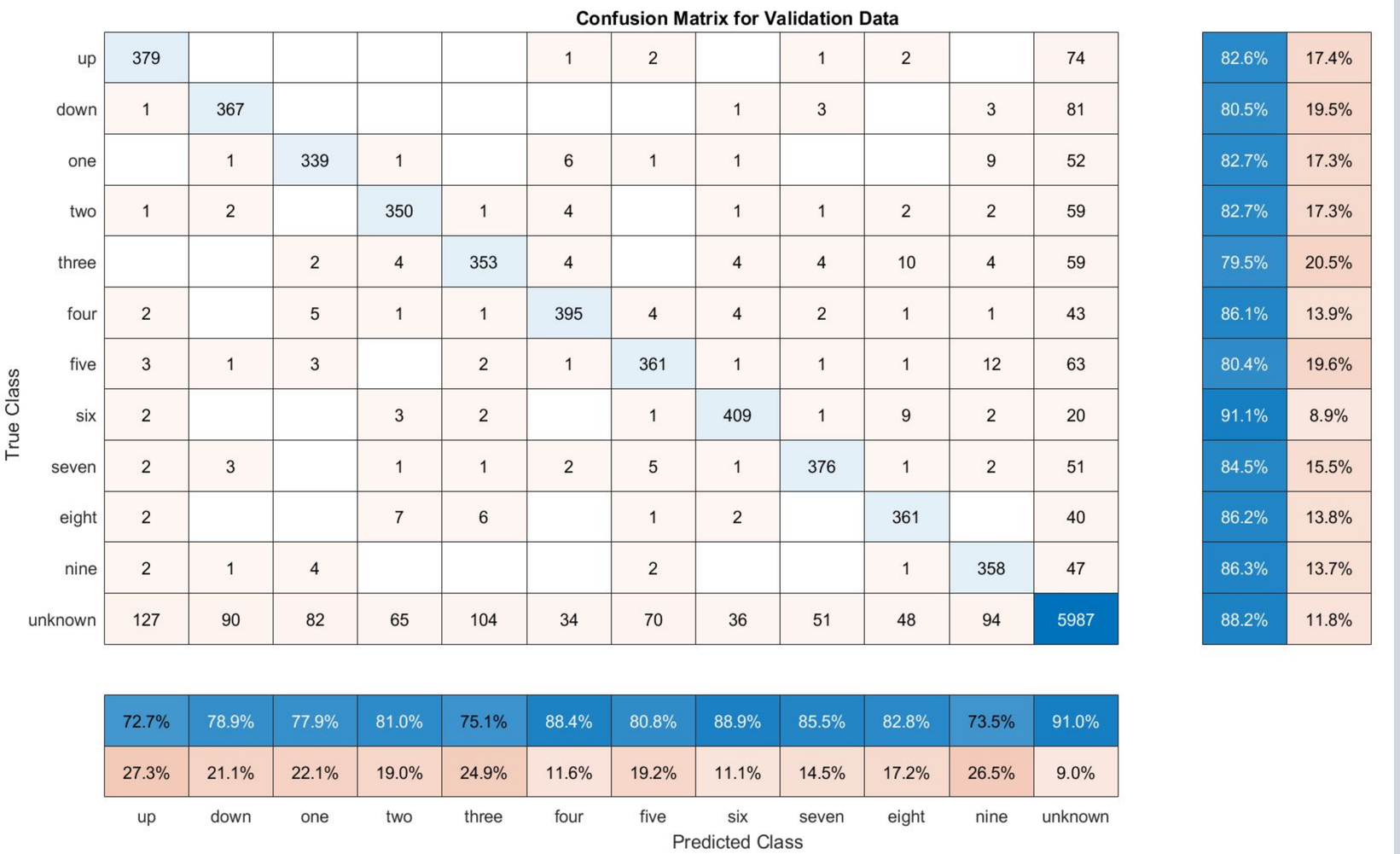


Fig. 4: Confusion matrix of a CNN trained with the full dataset of Mel spectrograms identifying keywords from the validation set
Validation Error: 13.6105%

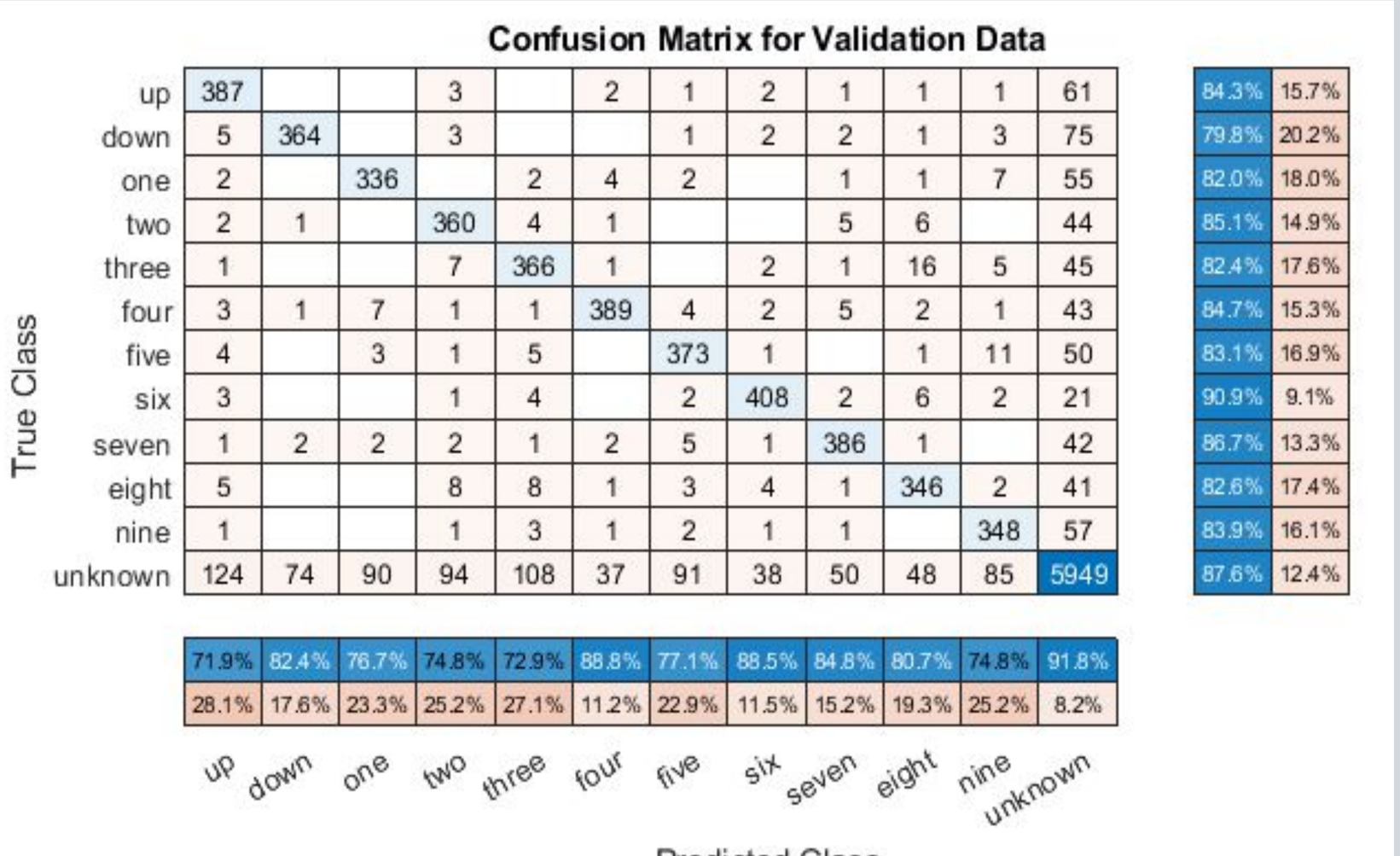


Fig. 5: Confusion matrix of a CNN trained with the full dataset of Mel spectrograms of both original and amplitude modulated signals identifying keywords from the validation set
Validation Error: 13.8085%

SULTS

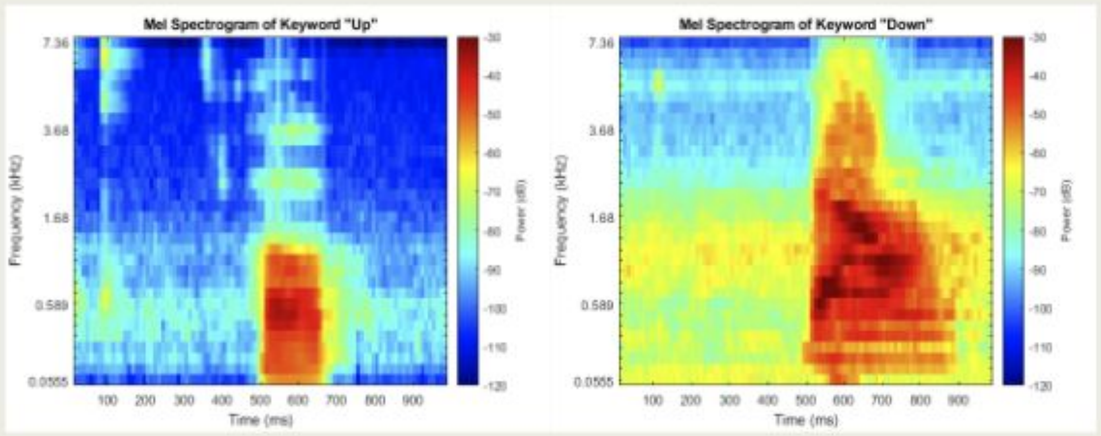


Fig. 3: The left is a Mel spectrogram and recording of the word "up" and the right is a Mel spectrogram and recording of the word "down".

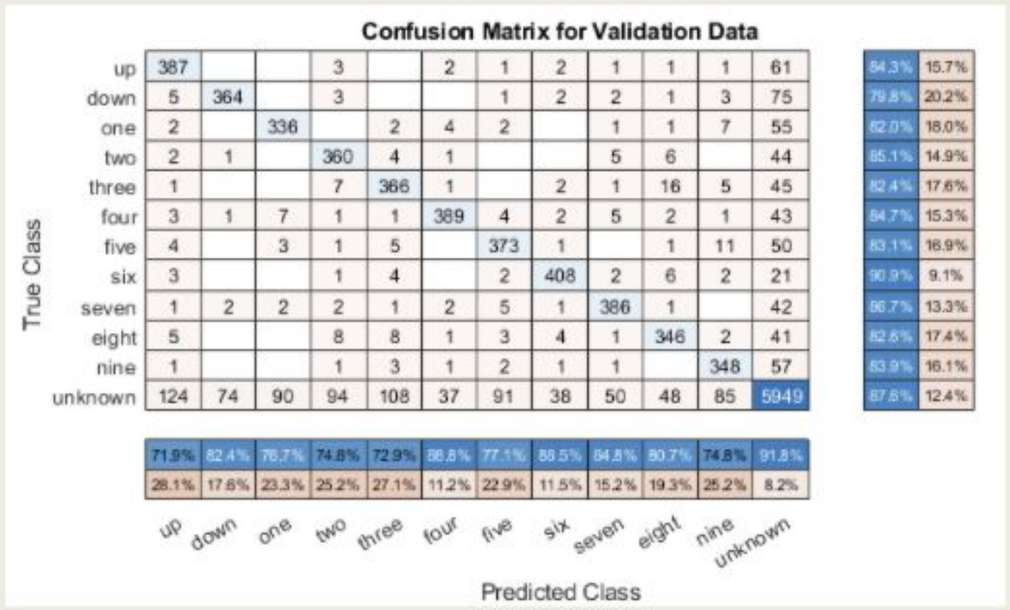


Fig. 5: Confusion matrix of a CNN trained with the full dataset of Mel spectrograms of both original and amplitude modulated signals identifying keywords from the validation set
Validation Error: 13.8085%

CONCLUSIONS

- It is feasible to perform keywords spotting using a convolutional neural network for volume adjustments of hearing devices.
- Mel-based spectral information is vital to train a neural network for keywords spotting, but combining this with amplitude information is a viable method
- Results are consistent with previous studies which show that with 8 or more bands, amplitude modulated audio can be recognized by human subjects.
- More research into amplitude modulation is necessary to determine if solely amplitude modulated signal processing is feasible for keywords spotting systems

REFERENCES

[1] Y. Zhang, N. Suda, L. Lai, and V. Chandra, "Hello Edge: Keyword Spotting on Microcontrollers," *ArXiv*, vol. abs/1711.07128, 2017.

[2] J. H. Won, C. Lorenzi, K. Nie, X. Li, E. M. Jameyson, W. R. Drennan, and J. T. Rubinstein, "The ability of cochlear implant users to use temporal envelope cues recovered from speech frequency modulation," *The Journal of the Acoustical Society of America*, vol. 132, no. 2, pp. 1113–1119, 2012.

[3] E. Scheirer and M. Slaney, "Construction and evaluation of a robust multifeature speech/music discriminator," 1997 IEEE International Conference on Acoustics, Speech, and Signal Processing, 1997, pp. 1331-1334.

[4] P. Warden, "Launching the Speech Commands Dataset," Google AI Blog, 24-Aug-2017. [Online]. Available: https://ai.googleblog.com/2017/08/launching-speech-commands-dataset.html.

ACKNOWLEDGEMENTS

This work was supported by the National Science Foundation REU Award #1757395 at the University of Washington Bothell.