Classifying dolphin whistles using convolutional neural networks

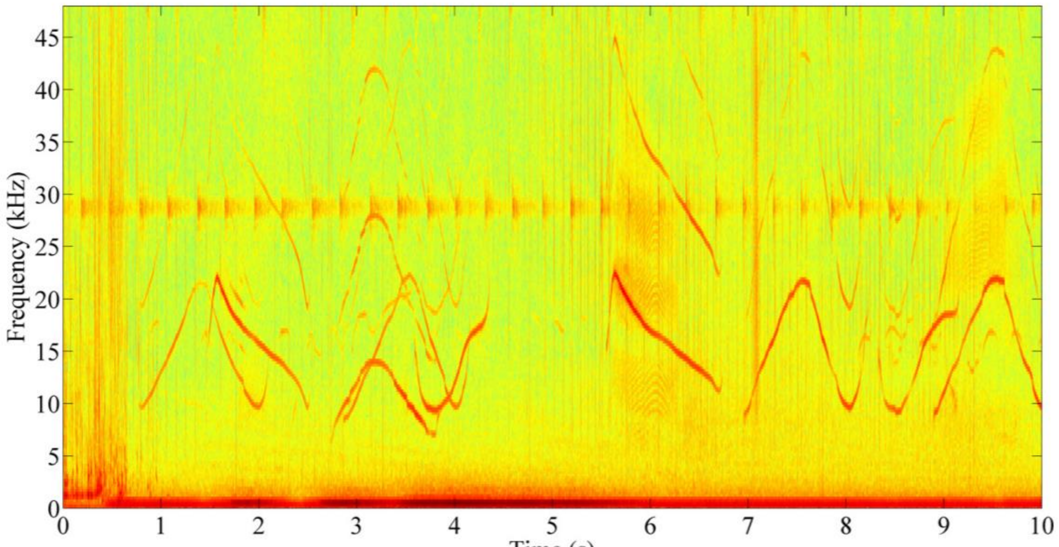
Genevieve Flaspohler, Tammy Silva, Aran Mooney, Yogesh Girdhar

Massachusetts Institute of Technology and
the Woods Hole Oceanographic Institute Joint Program

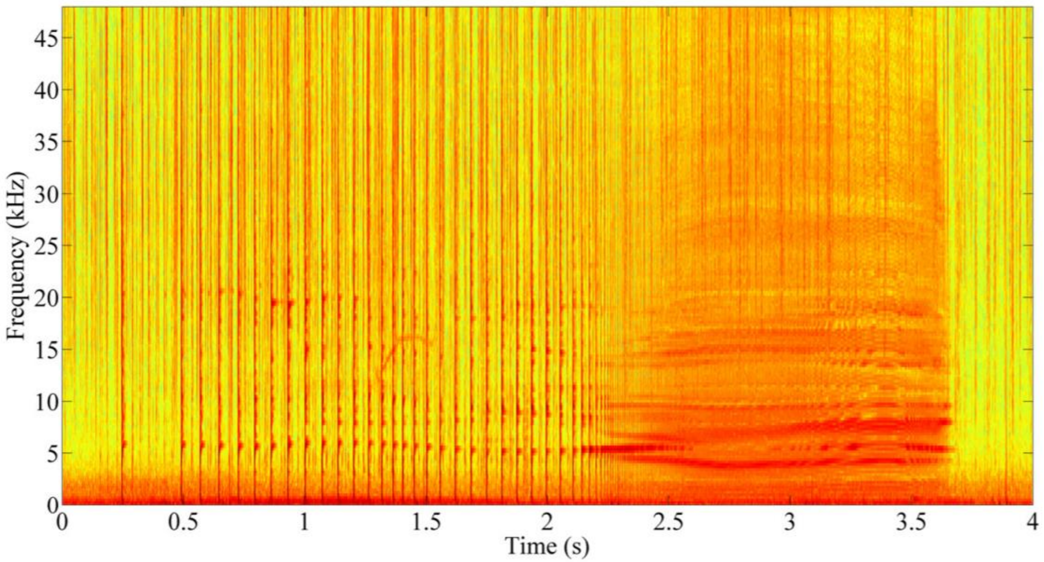
Acoustics ’17 Boston - 28 June 2017

WOODS HOLE OCEANOGRAPHIC INSTITUTION 1930

WARPLab



(a)



(b)

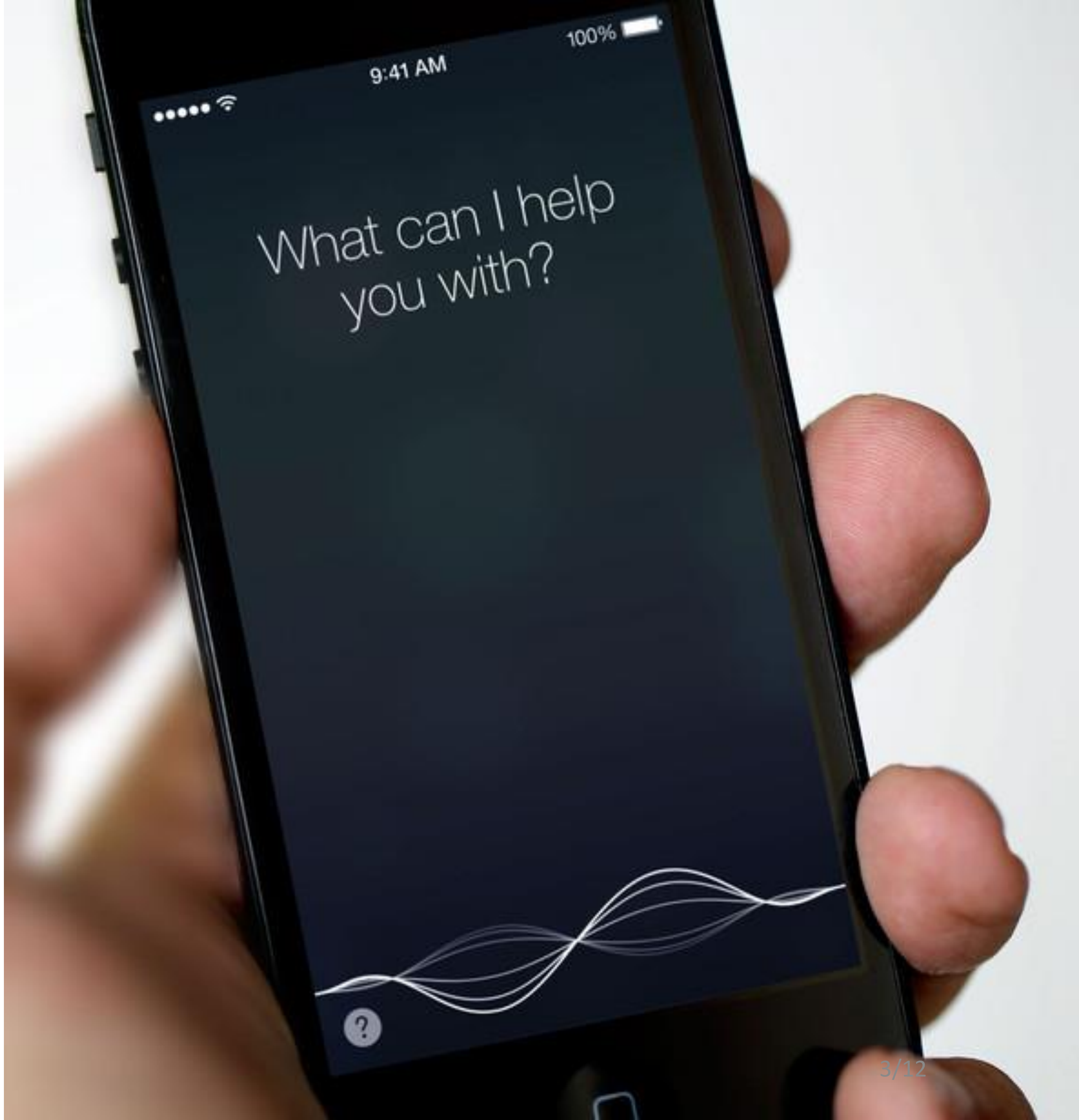
Machine learning

Dolphin species, type of sound, category of whistle, etc.

Silva, Tammy L., et al. ”Whistle characteristics and daytime dive behavior in pantropical spotted dolphins (Stenella attenuata) in Hawai’i measured using digital acoustic recording tags (DTAGs).” The Journal of the Acoustical Society of America 140.1 (2016): 421-429.

Human speech recognition systems like Siri and Google Now can achieve ~95% accuracy **using neural network-based recognition**.

https://www.androidheadlines.com/2017/06/google-improves-voice-recognition-hits-95-accuracy.html



Hypothesis: Neural network methods from Automatic Speech Recognition (ASR) can be adapted to classify dolphin whistles.

Machine learning as function approximation

$$f : input \rightarrow category$$

$f($  $) = dog$

$f($  $) = cat$

How do we come up with $\boldsymbol{f}(\cdot)$?

1. Task-specific features, learn $f(\cdot)$
   (Previous work)
2. Standard features, learn $f(\cdot)$
   (SVMs, Feed-forward NN, etc.)
3. Learn features and $f(\cdot)$ directly
   (Convolutional neural methods)

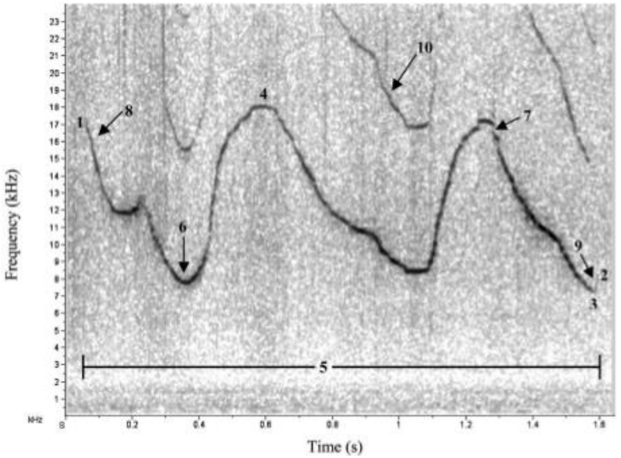
Task-specific Features:

Based on 10 features*:

1) Start frequency
2) End frequency
3) Minimum frequency
4) Maximum frequency

…

*Oswald, Julie N., et al. (2007) [3]

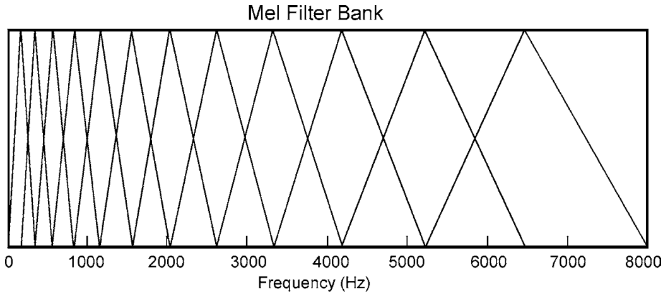


MFCC Features*:

A 26-dimensional vector that represents the **power in each frequency band**

Filter bank designed to **emulate human hearing** physiology

*Mermelstein, Paul. (1976) [6]

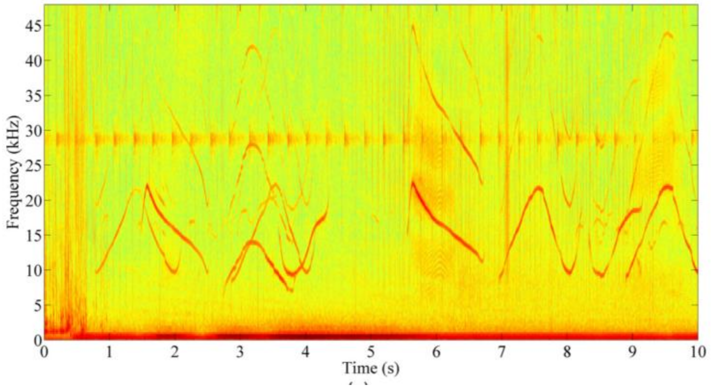


Full Spectrogram data:

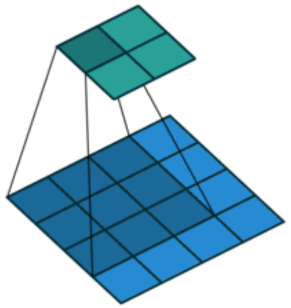
Calculated using a sliding window FFT

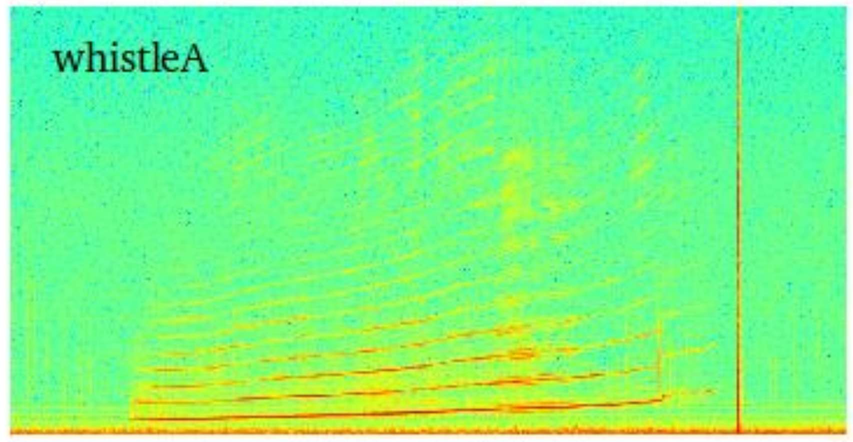
1024-pt FFT on sliding Hanning window with 50% overlap.

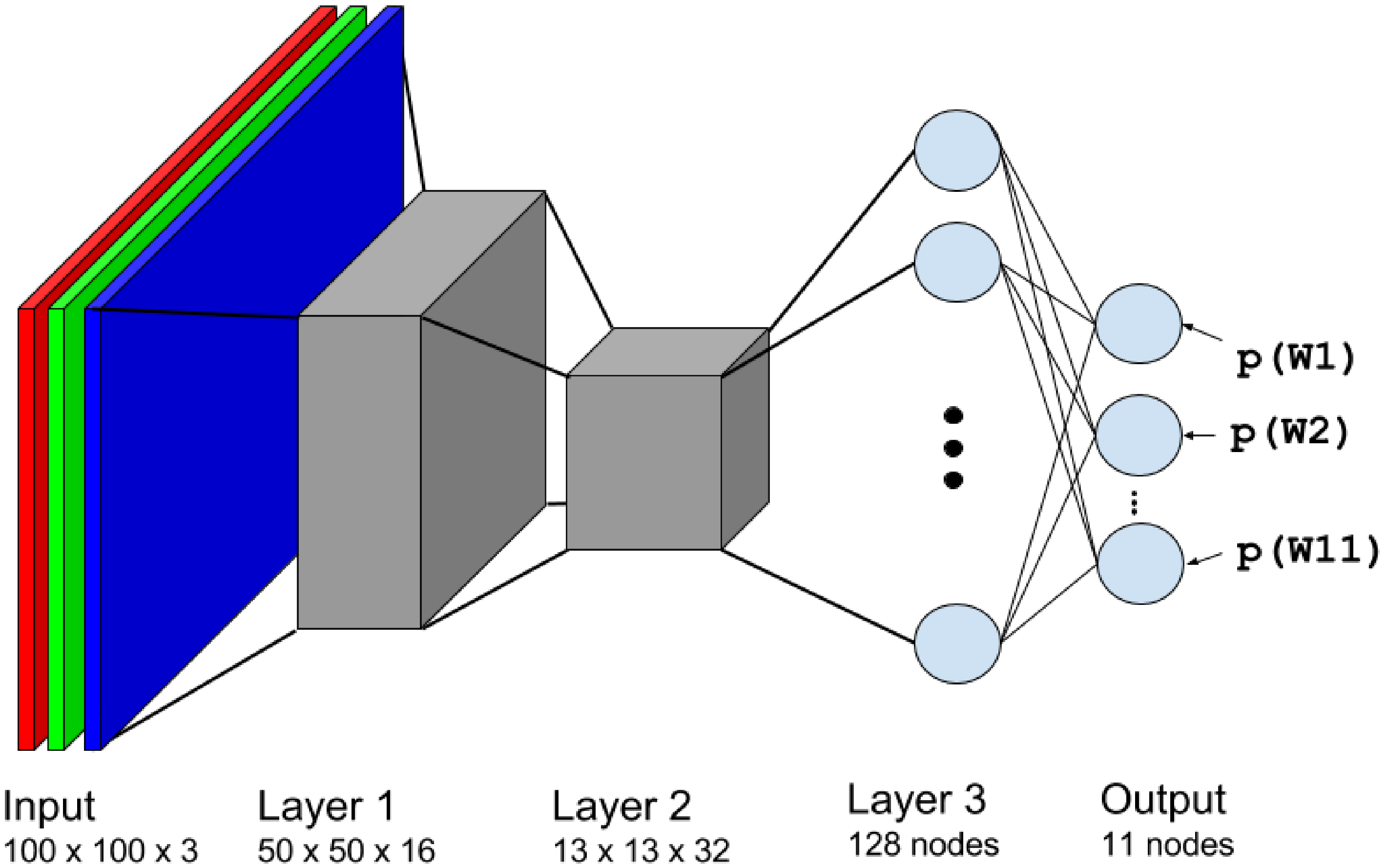
Contains "all" data available in signal



Convolutional neural network (CNN)



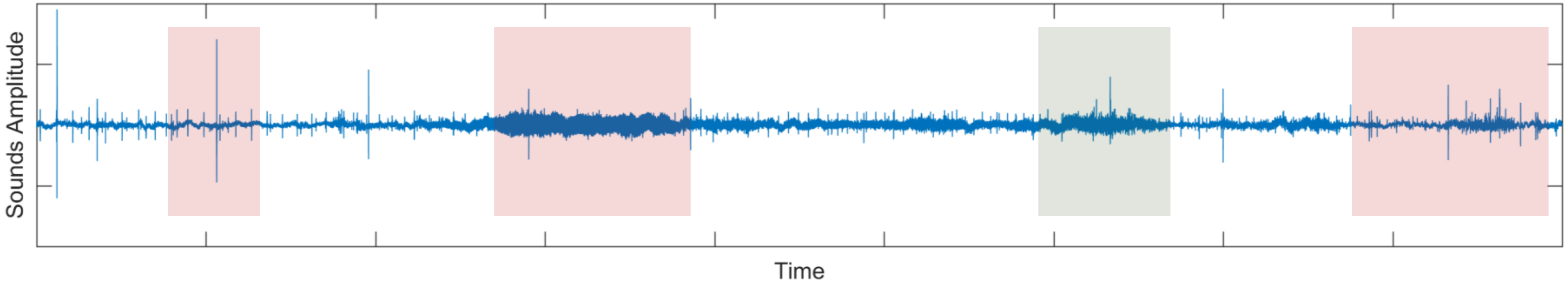




Machine learning dataset

Day 1

Day 2

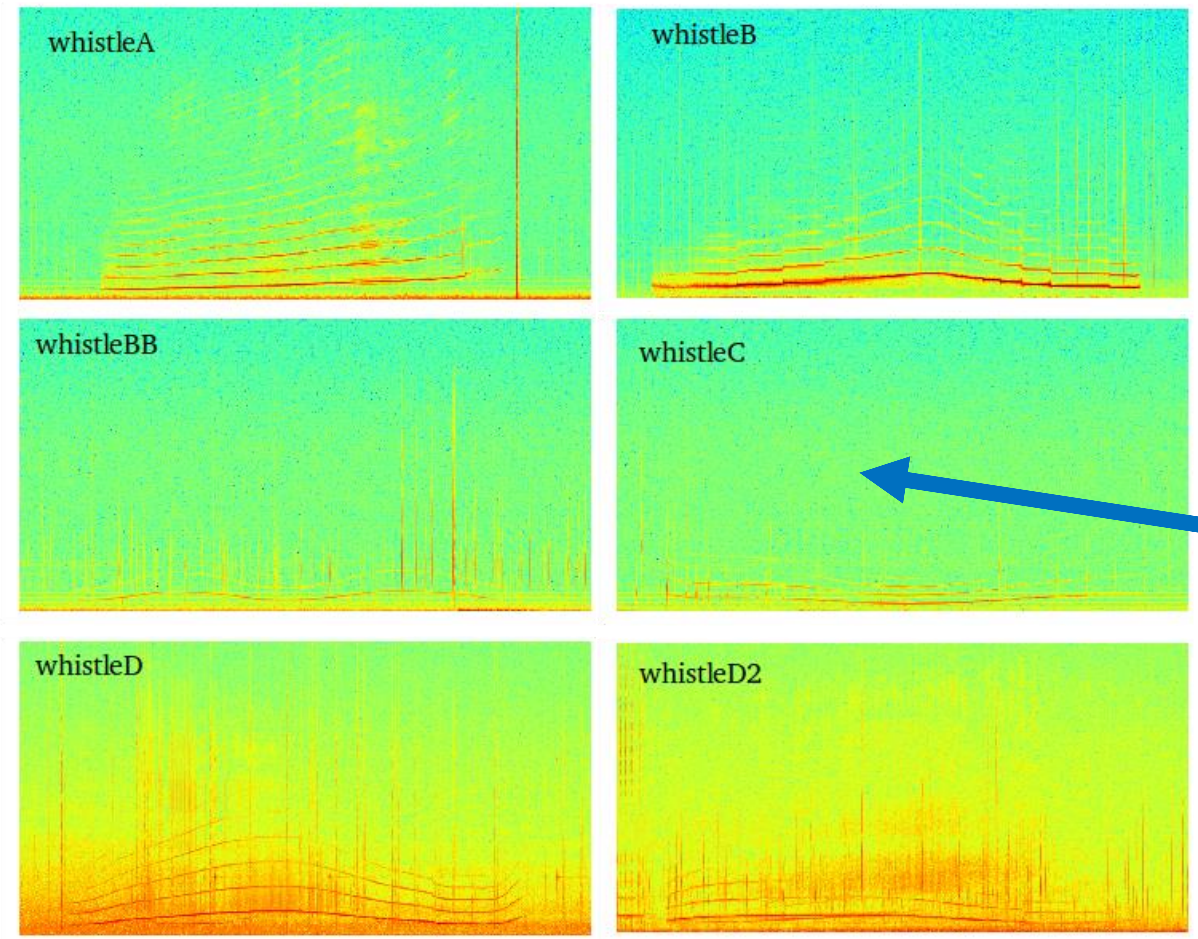


8,632 total vocalizations

339 “clear” whistles

Data generously provided by Tammy L. Silva ad T. Aran Mooney [1]

Machine learning dataset



339 total whistle clips

Classified into 11 categories

x 10

Random shuffle

80% training clips

20% testing

CNNs attain maximum classification accuracy

Linear discriminant analysis	17.2% in 11-way classification*
Decision trees	20.6%*
Linear discriminant analysis	59.7%
Decision trees	61.5%
SVM polynomial kernel	82.8%
Feed forward NN	83.7%
Logistic regression	76.8%
Convolutional NN	**85.9% average accuracy**

Task specific

MFCC

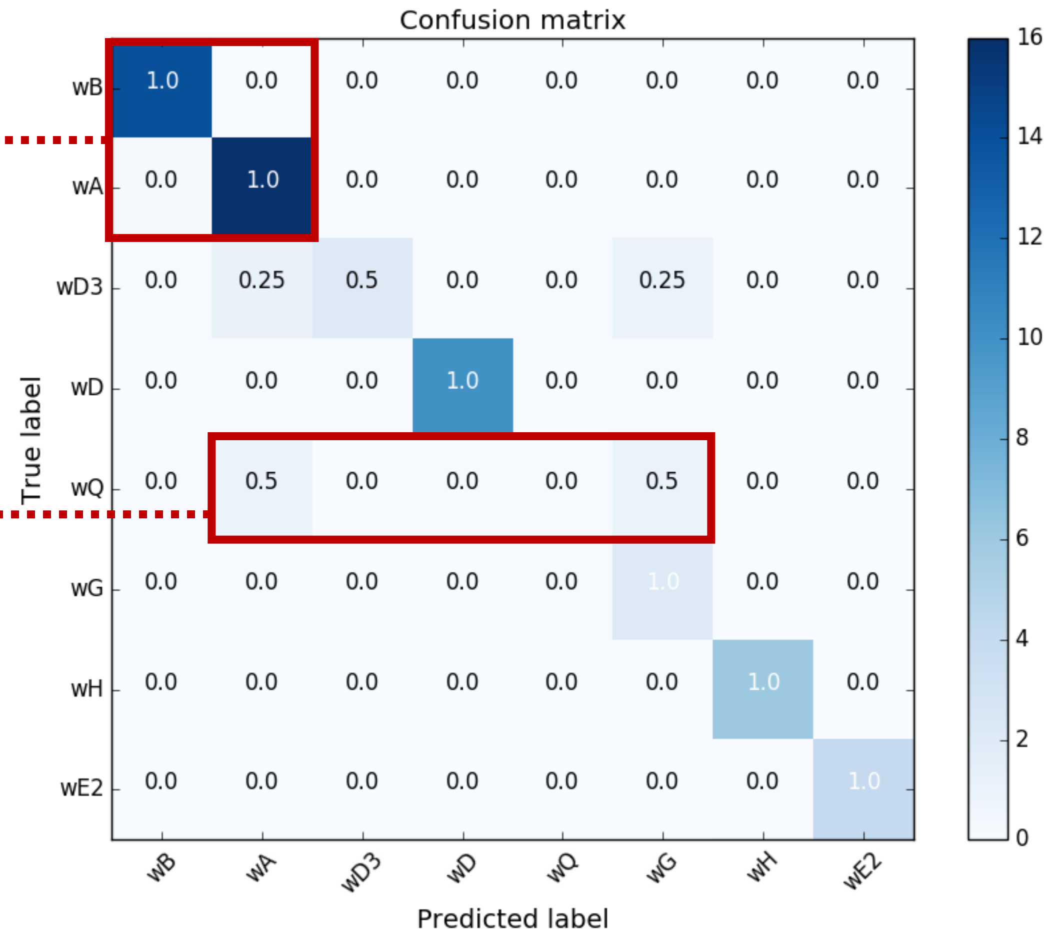
Spectrogram

*vs 33.5 and 33.6 reported for 8-way classification in Oswald, Julie N., et al. (2007)

CNN performance is label dependent

High classification accuracy in classes with many examples

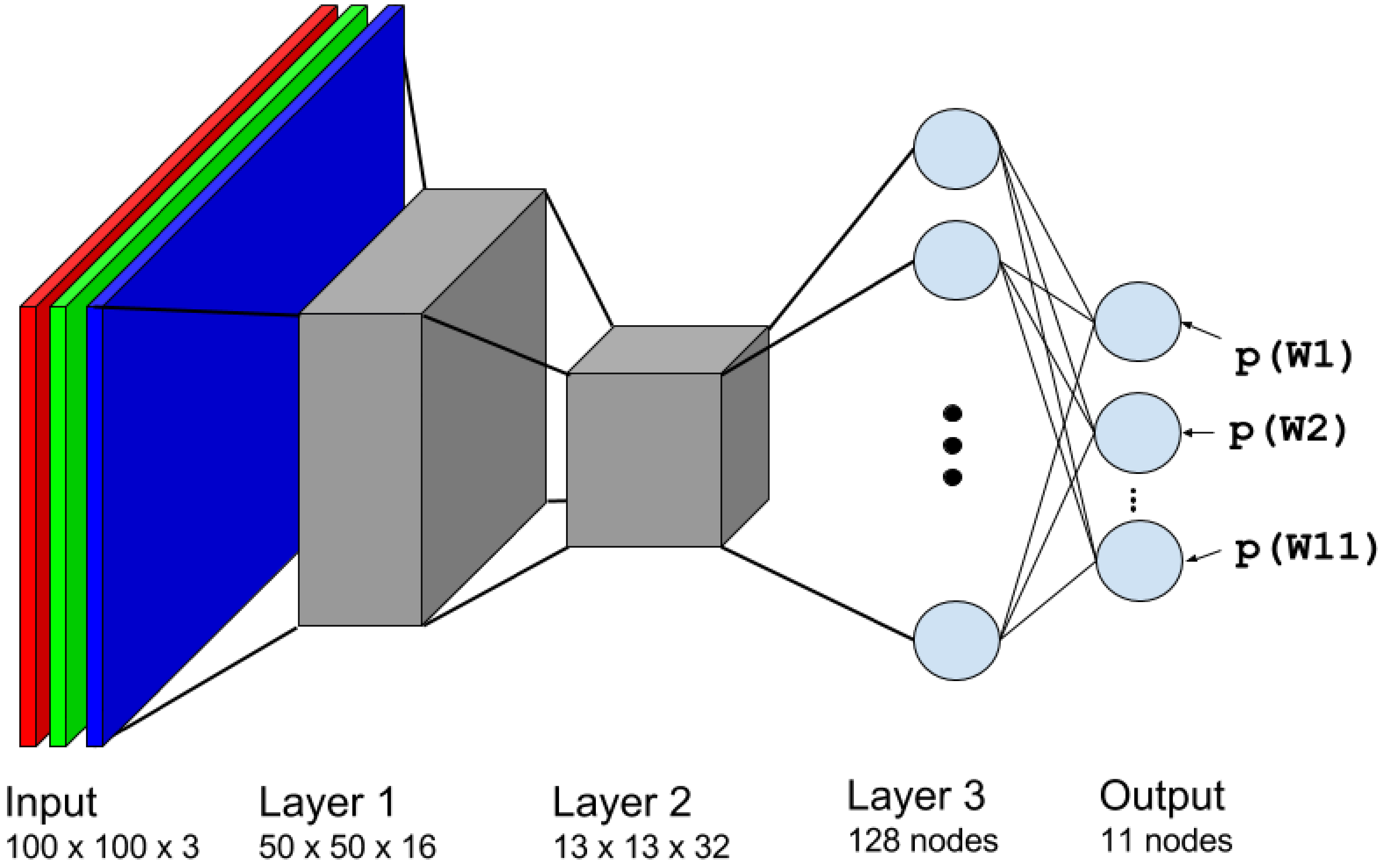
Low accuracy for classes with few examples



Performance should continue to improve with additional training examples

Code: https://gitlab.com/warplab/dolphin-lang

Website: https://warp.whoi.edu



References

[1] Johnson, Mark P., and Peter L. Tyack. ”A digital acoustic recording tag for measuring the response of wild marine mammals to sound.” IEEE journal of oceanic engineering 28.1 (2003): 3-12.

[2] Silva, Tammy L., et al. ”Whistle characteristics and daytime dive behavior in pantropical spotted dolphins (Stenella attenuata) in Hawai i measured using digital acoustic recording tags (DTAGs).” The Journal of the Acoustical Society of America 140.1 (2016): 421-429.

[3] Oswald, Julie N., et al. ”A tool for real-time acoustic species identification of delphinid whistles.” The Journal of the Acoustical Society of America 122.1 (2007): 587-595.

[4] O’Shaughnessy, Douglas. ”Invited paper: Automatic speech recognition: History, methods and challenges.” Pattern Recognition 41.10 (2008): 2965-2979

[5] Abdel-Hamid, Ossama, Li Deng, and Dong Yu. "Exploring convolutional neural network structures and optimization techniques for speech recognition." *Interspeech*. 2013.

[6] Mermelstein, Paul. "Distance measures for speech recognition, psychological and instrumental." *Pattern recognition and artificial intelligence* 116 (1976): 374-388.


AB-2
IF FOUND ADRIFT CONTACT
MOORING OPERATIONS GROUP
WOODS HOLE OCEANOGRAPHIC
INSTITUTION
WOODS HOLE MA 02543 USA

Feed-forward neural network