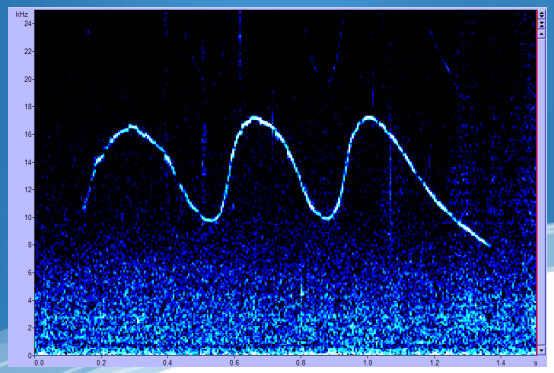
A comparison of automated detectors for tonal signals from delphinids

Julie Oswald, Kerry Dunleavy, Shannon Coates, Marie A. Roch, Doug Gillespie, David K. Mellinger

Introduction

- Delphinid whistles
  - Long duration, tonal sounds
  - Frequency and amplitude modulated
  - Start and end points often undefined
- Automated detection and contour extraction can be challenging
  - Several different approaches



Bio·Waves Incorporated

Objectives

- **Evaluate and compare performance** of three automated whistle detection and contour extraction algorithms
  - ***Silbido*** (Roch et al.[1])
  - **Ishmael's** tonal detector (Mellinger et al.[2])
  - **PAMGuard's** whistle and moan detector (Gillespie et al.[3])

[1] Roch et al. 2011, JASA 130:2212-2223

[2] Mellinger et al. 2011, JASA 129:4055-4061

[3] Gillespie et al. 2013, JASA 134:2427-2437

Bio·Waves Incorporated

Dataset

- Manually annotated whistles[1]
  - *Silbido* tracing tool
  - www.mobysound.org
- Four encounters per species
  - Different SNRs and overlapping sounds

Species	Whistle characteristics	n
Melon-headed whale	Low frequency, short duration	2,253
Bottlenose dolphin	Large frequency range, long duration	1,637
Short-beaked common dolphin	High frequency, short duration, overlapping whistles	3,505

[1] Roch et al. 2011, JASA 130:2212-2223

Detector Evaluation

- Each detector was run on the same wav files and results compared to manual annotations
  - *Silbido* Scoring Tool
- Different settings used for each detector
- Chose settings that provided best balance between precision and recall

Important!!

Results may have been different if different settings had been chosen

Bio·Waves Incorporated

Silbido Scoring Tool

- Compares whistle detector output to manual annotations
- Calculates five metrics:
  - Precision
  - Recall
  - Fragmentation
  - Percent coverage
  - Frequency deviation

Bio·Waves Incorporated

Precision

- Measure of ‘exactness’
  - Percentage of automated detections that were actually whistles
  - Aiming for 100%

$$Precision = \frac{TP}{TP + FP}$$

TP = True Positive detections
FP = False Positive detections

Bio·Waves Incorporated

Recall

- Measure of ‘completeness’
  - Percentage of whistles in a recording that were detected
  - Aiming for 100%

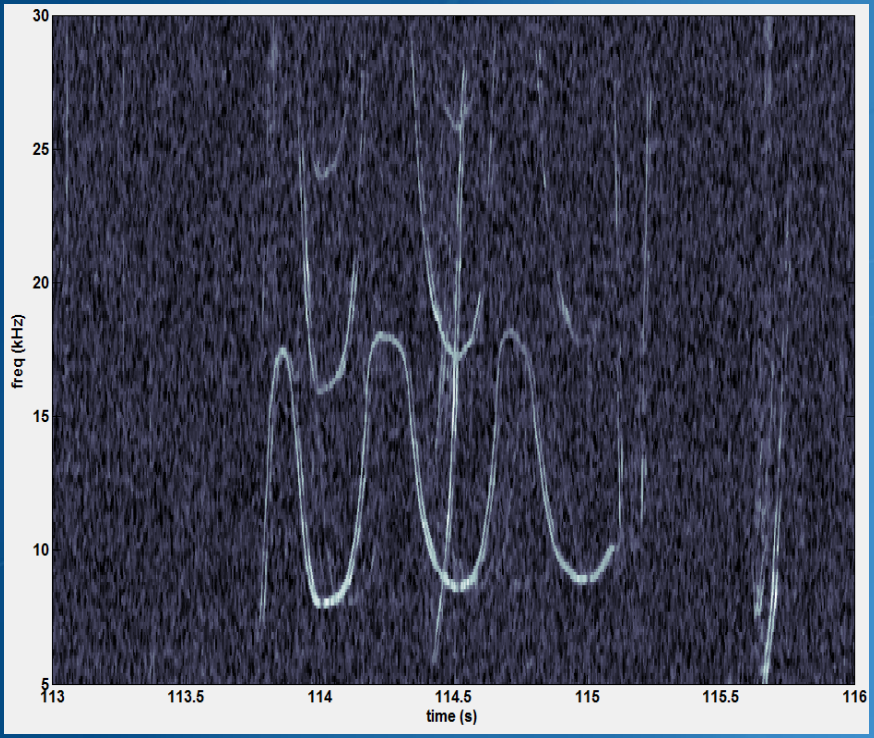
$$Recall = \frac{TP}{TP + FN}$$

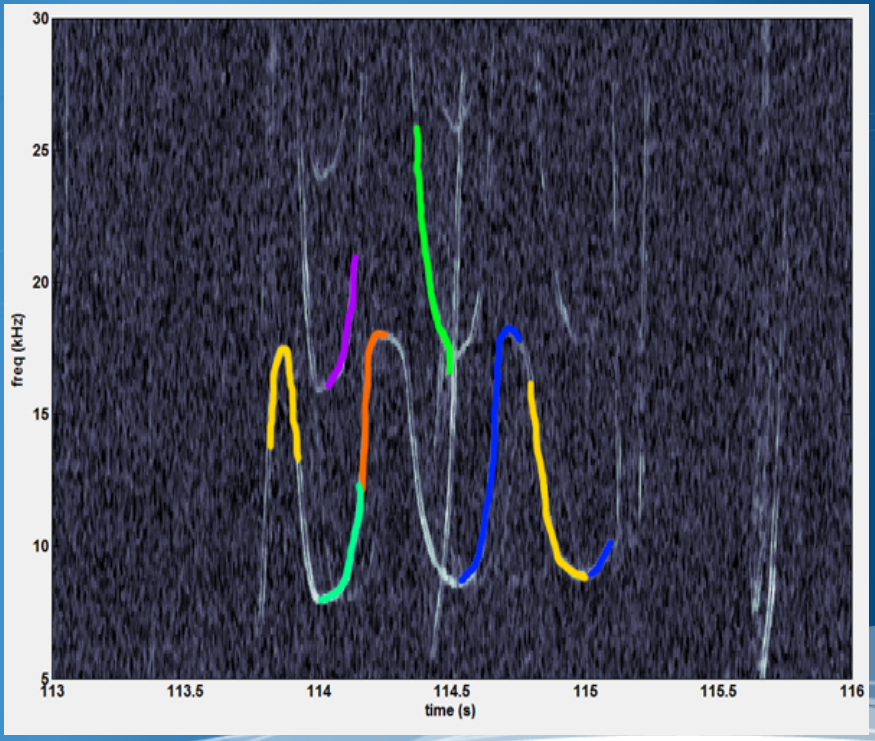
TP = True Positive detections
FP = False Negative (missed) detections

Bio·Waves Incorporated

Fragmentation

- Average number of detections per individual whistle
- Aiming for a value of 1

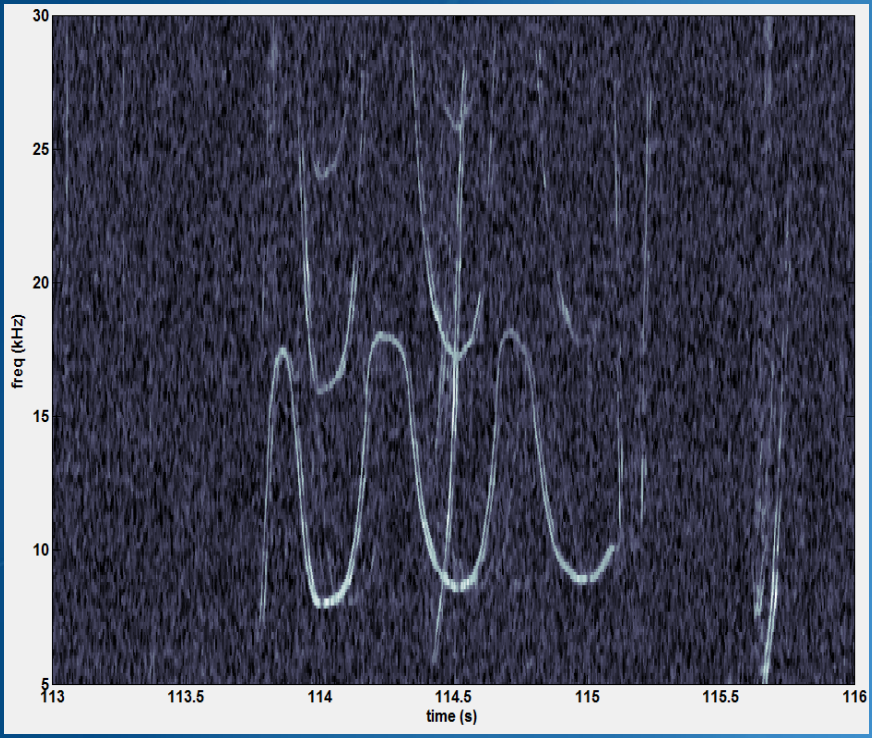


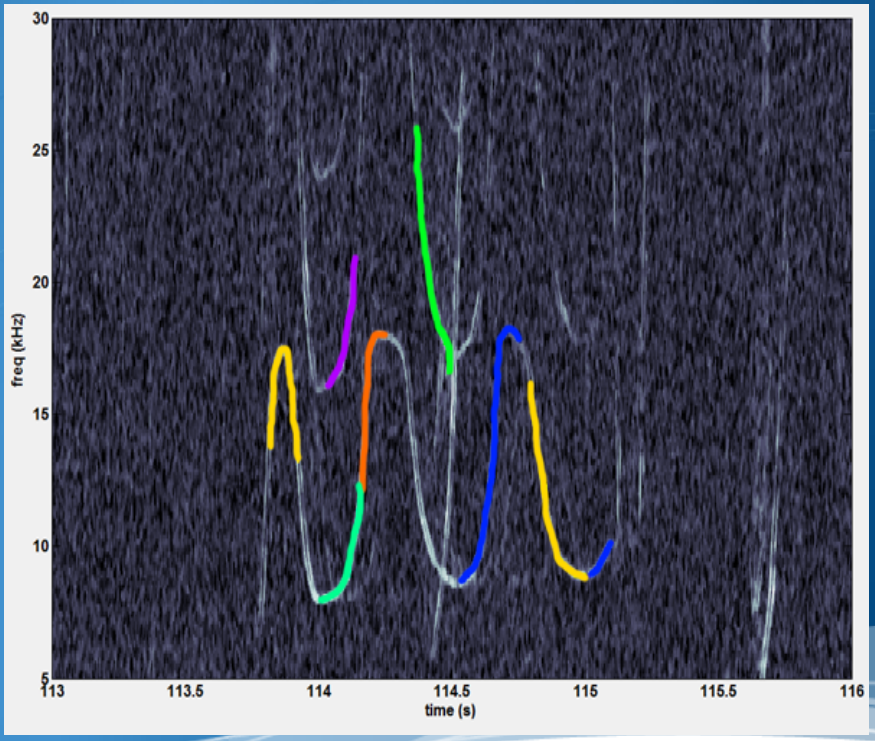


Bio·Waves Incorporated

Percent coverage

- Average percentage of the duration of individual whistle contours that were extracted
- Aiming for 100%

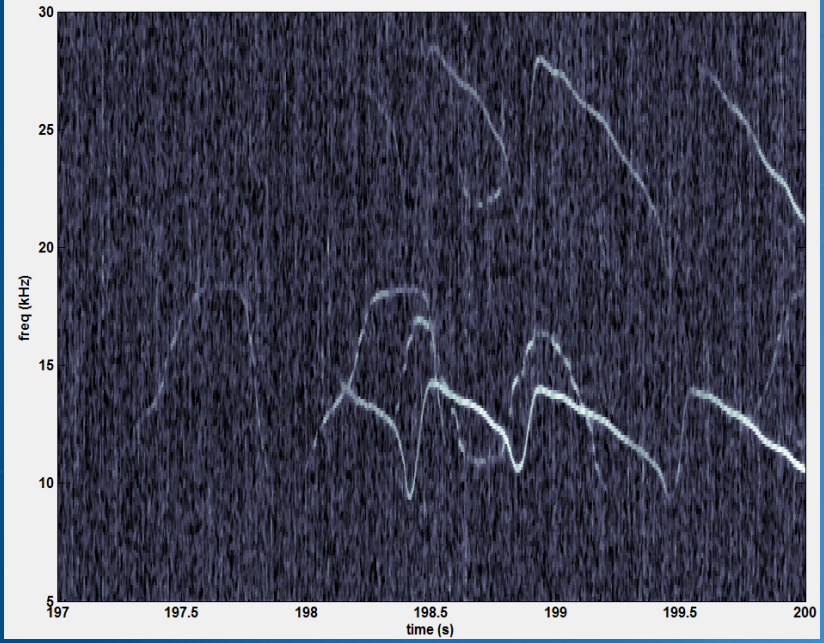


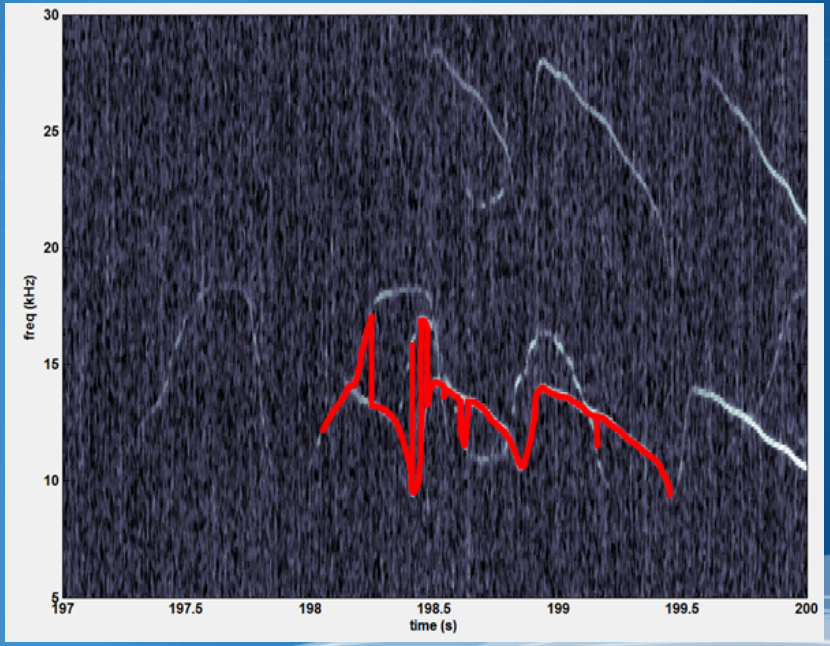


Bio·Waves Incorporated

Frequency deviation

- Absolute difference in frequency between extracted time-frequency contour and actual whistle
- Aiming for a value of 0



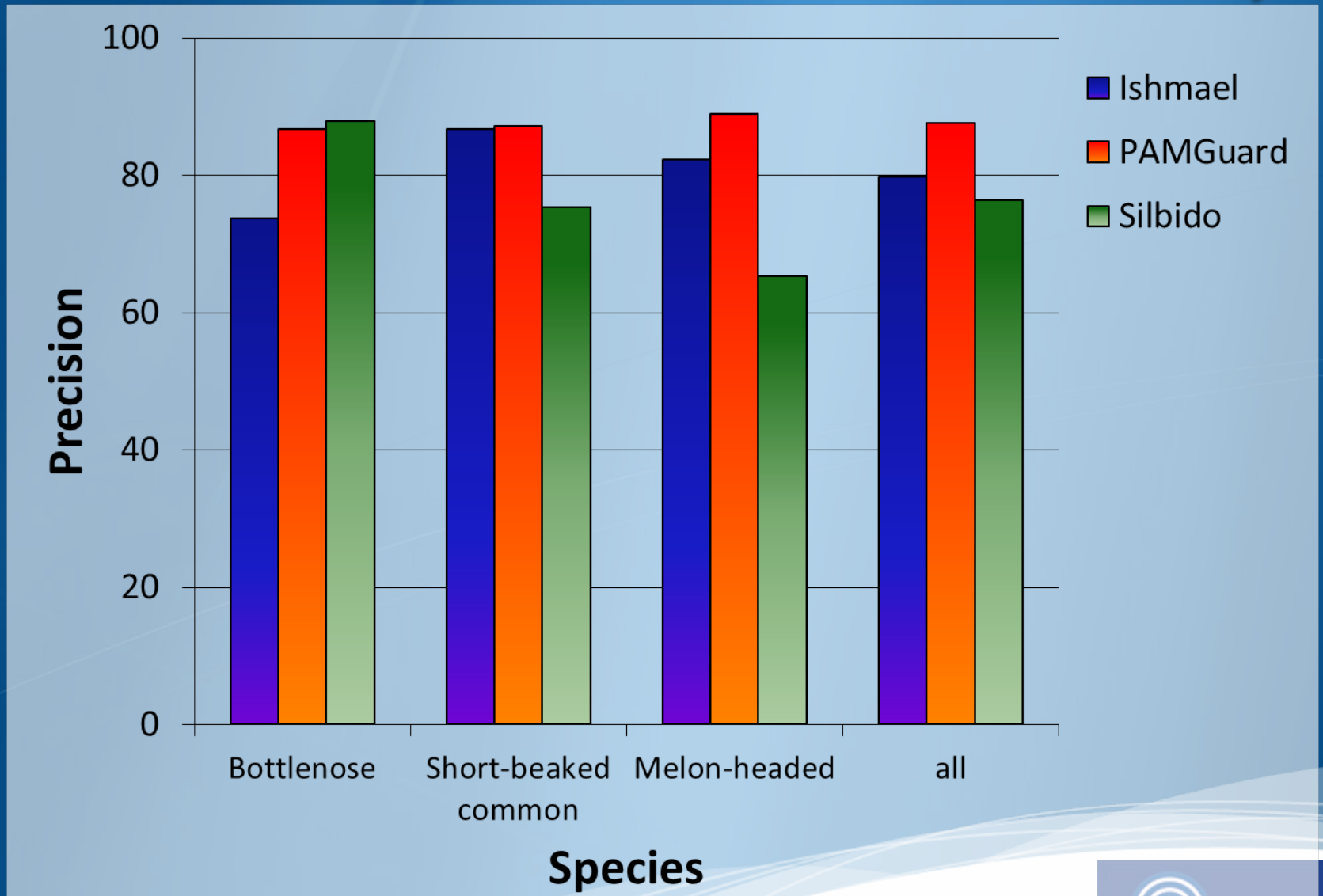


Bio·Waves Incorporated

Results

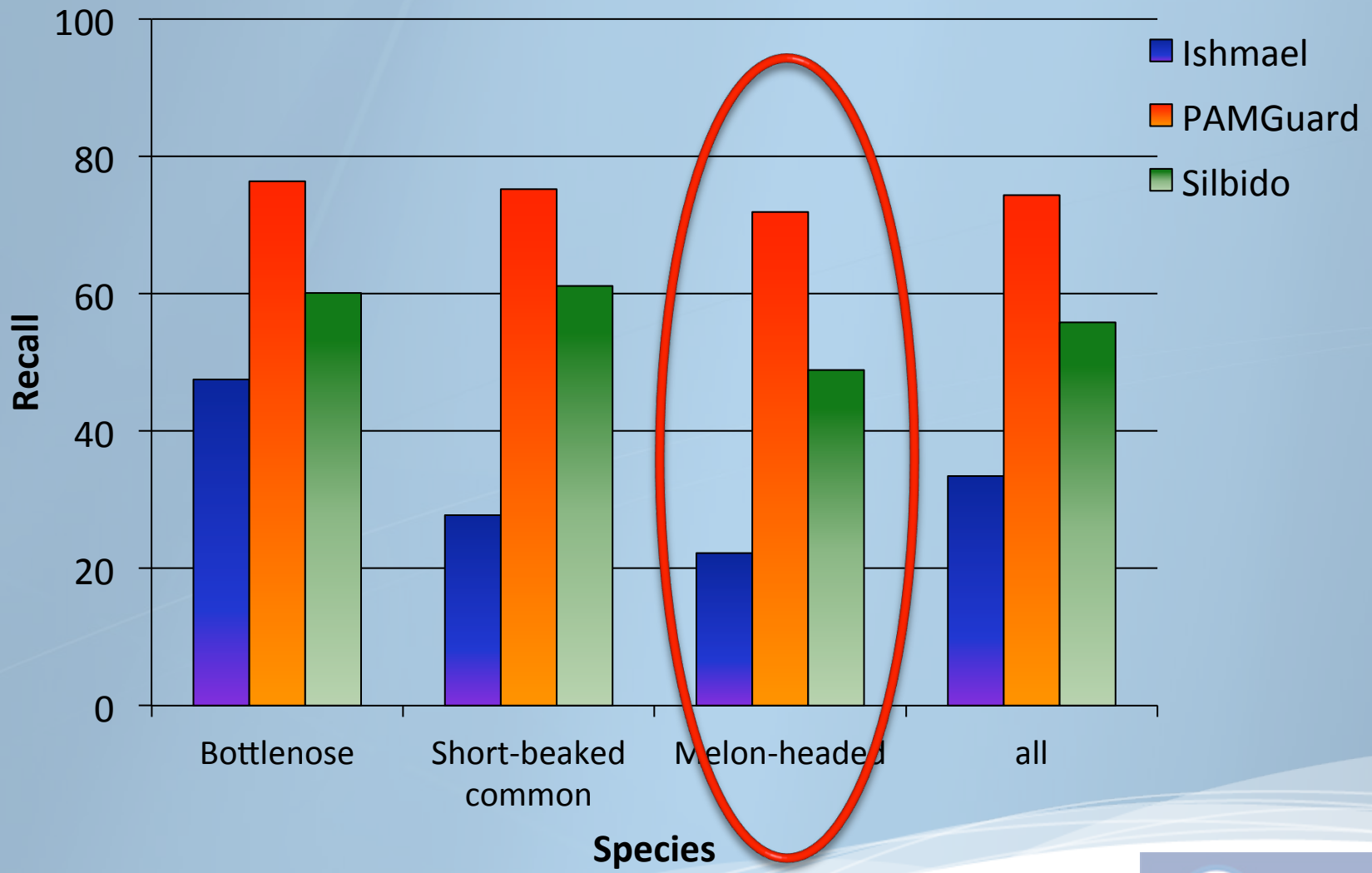
Bio·Waves Incorporated

Precision – measure of accuracy



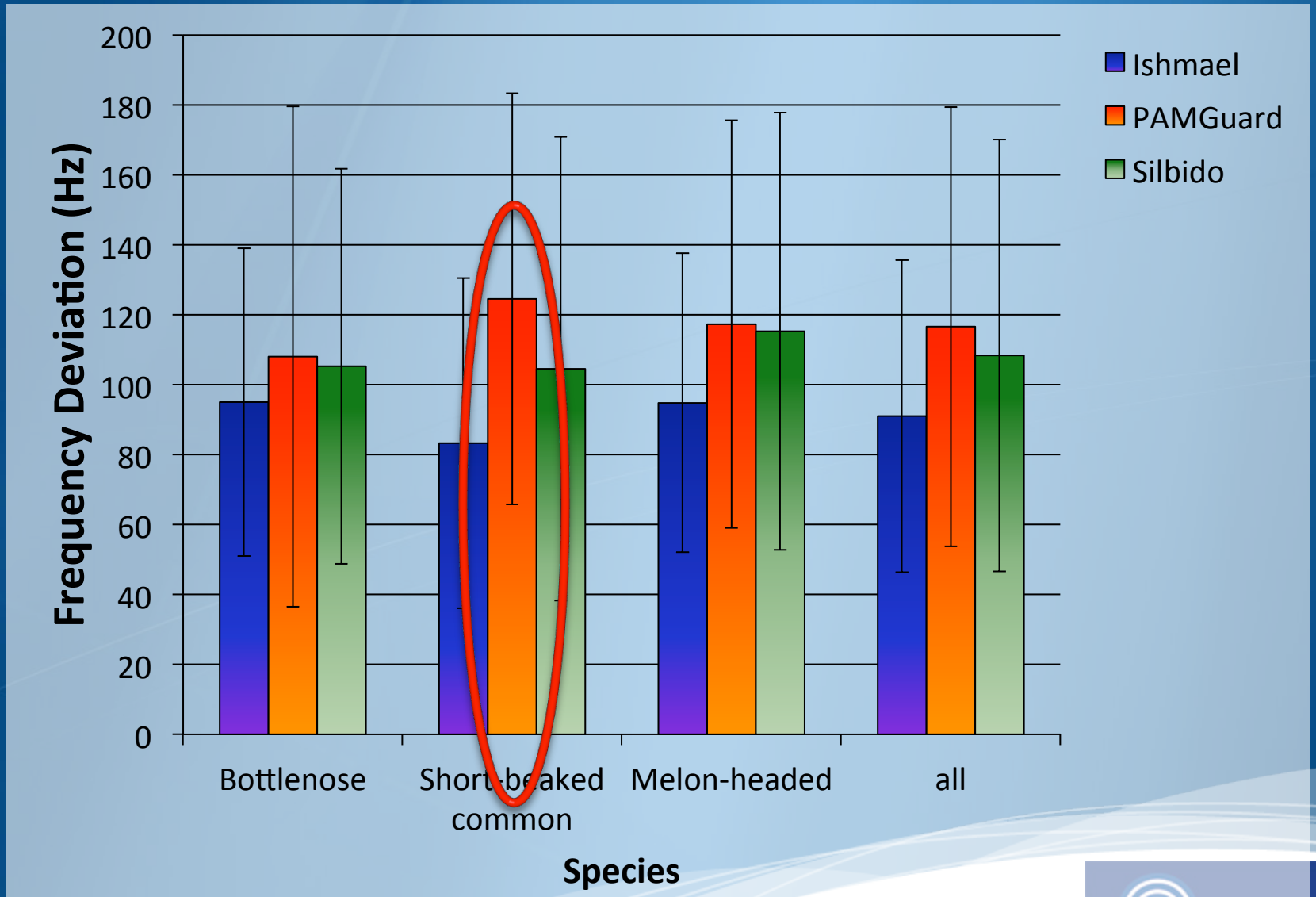
Bio·Waves Incorporated

Recall – measure of completeness



Bio·Waves Incorporated

Frequency deviation



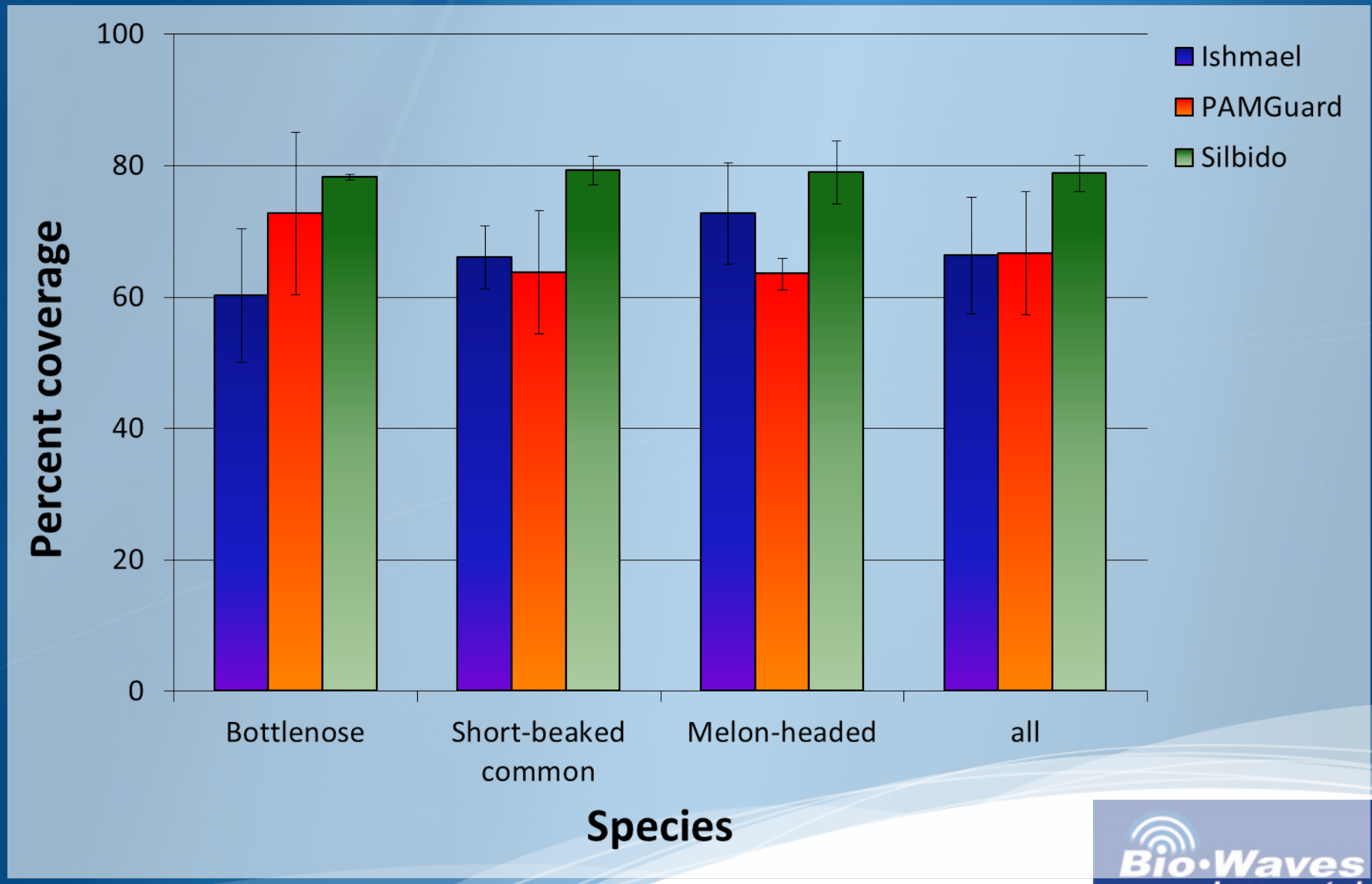
Bio·Waves Incorporated

Frequency deviation

- Driven by analysis window size (to an extent)
  - 8ms window, 192 kHz sampling rate = 125 Hz frequency resolution
- Highest deviation of 124 Hz is within this window size

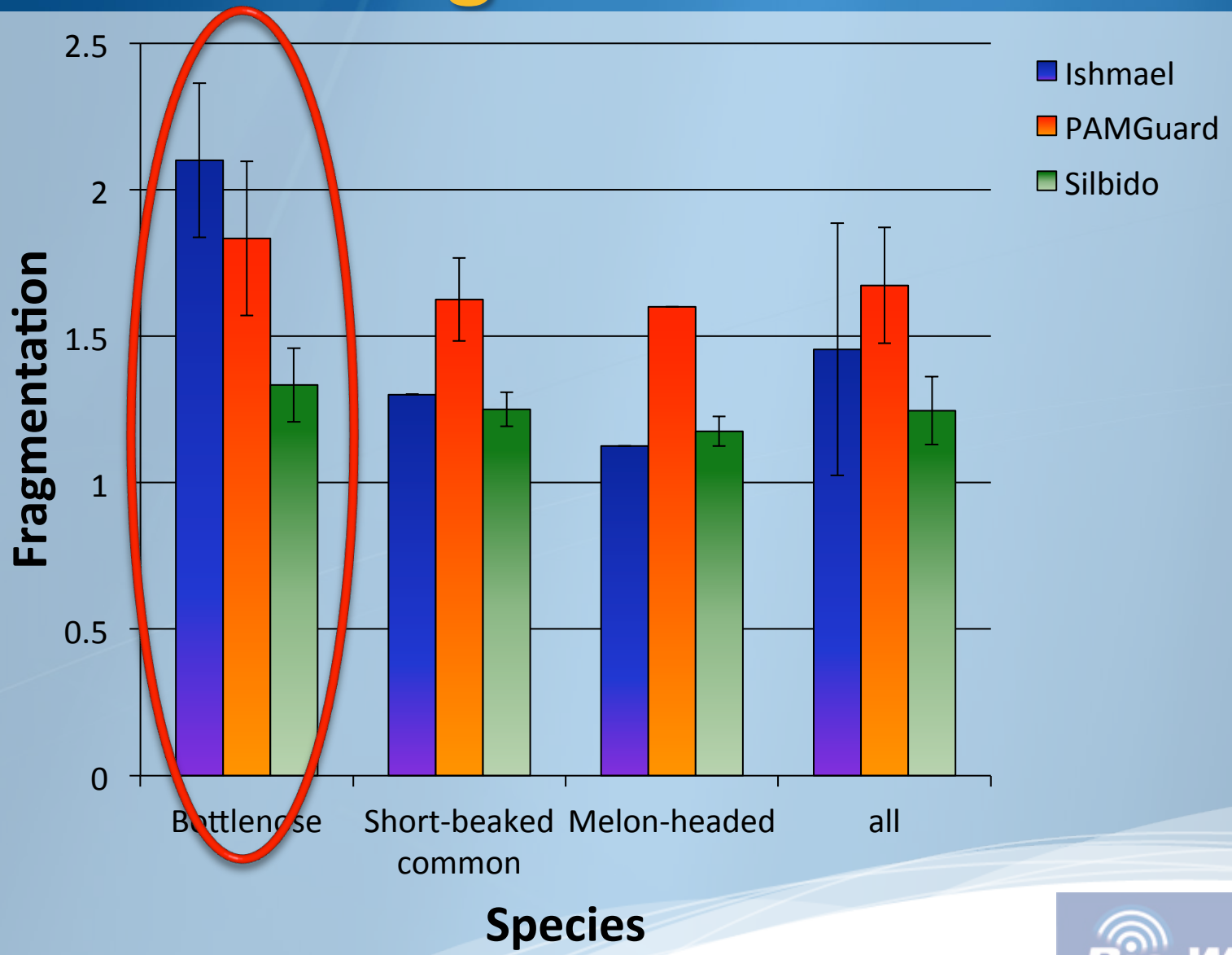
Bio·Waves Incorporated

Percent coverage



Bio·Waves Incorporated

Fragmentation



Bio·Waves Incorporated

Discussion

Bio·Waves Incorporated

Factors influencing detector performance

- Comparison metrics influenced by detector settings
- Examined recordings with different whistle characteristics and SNR
- All three detectors missed short whistles with low SNR
  - Likely reason for low recall scores for melon-headed whales
- Ishmael missed whistles with the lowest SNR

Bio·Waves Incorporated

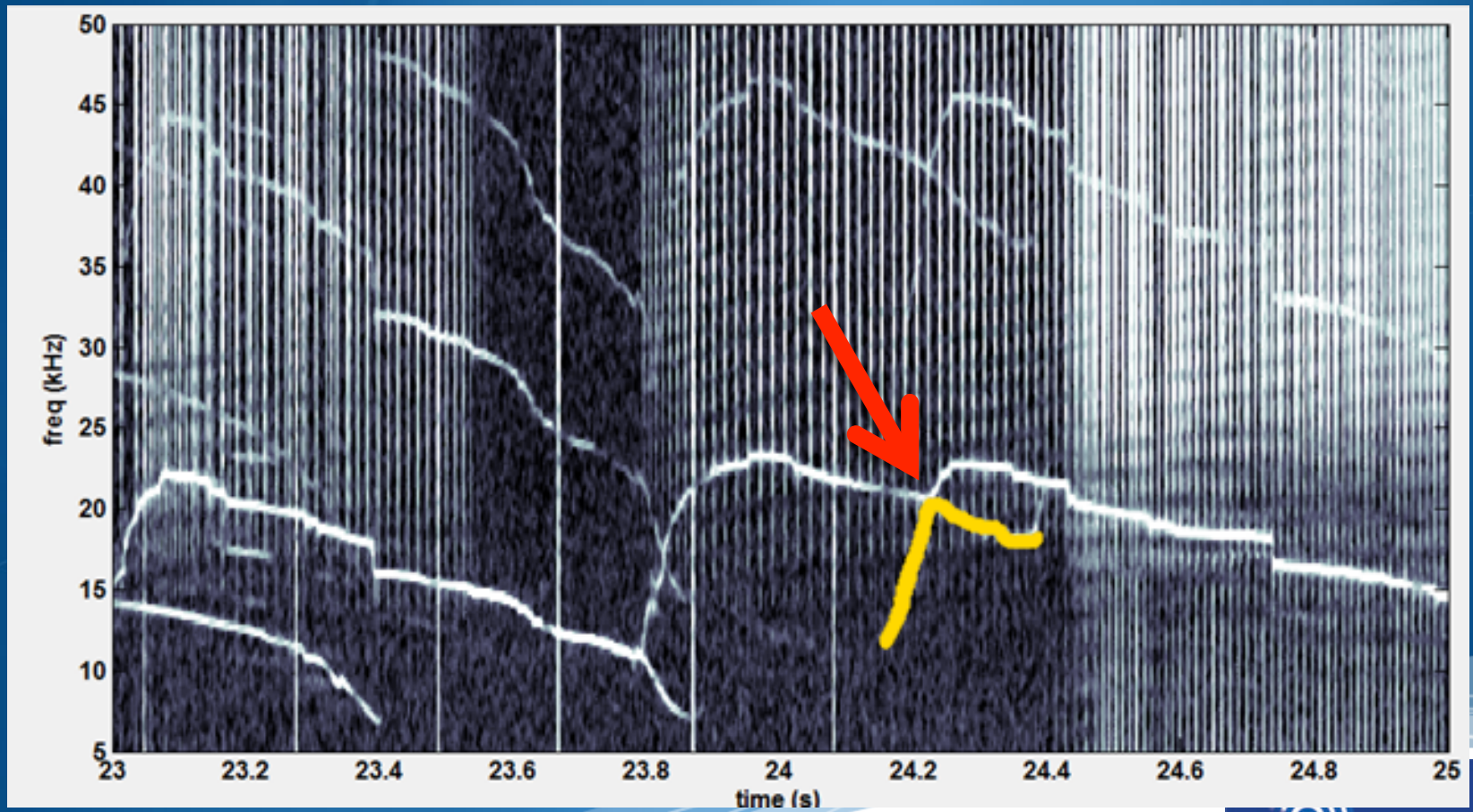
Factors influencing detector performance

- Other factors that affected performance of detectors
  - Presence of overlapping whistles
  - Presence of clicks and buzzes

Bio·Waves Incorporated

Overlapping whistles

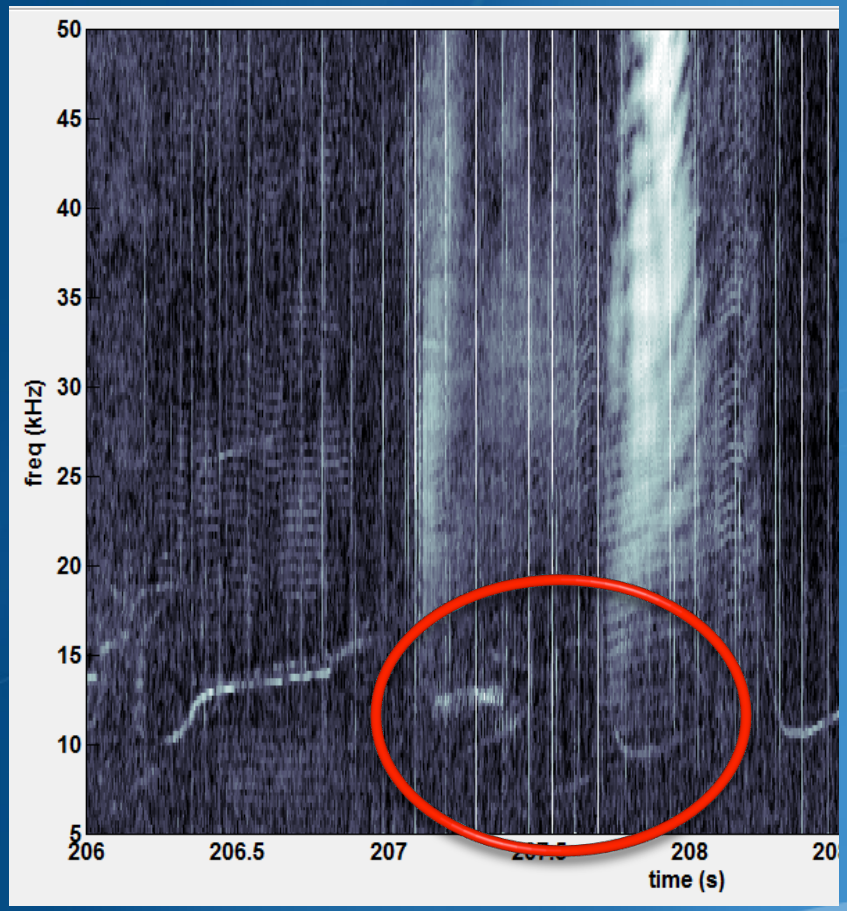
- Increased number of missed whistles and false positives

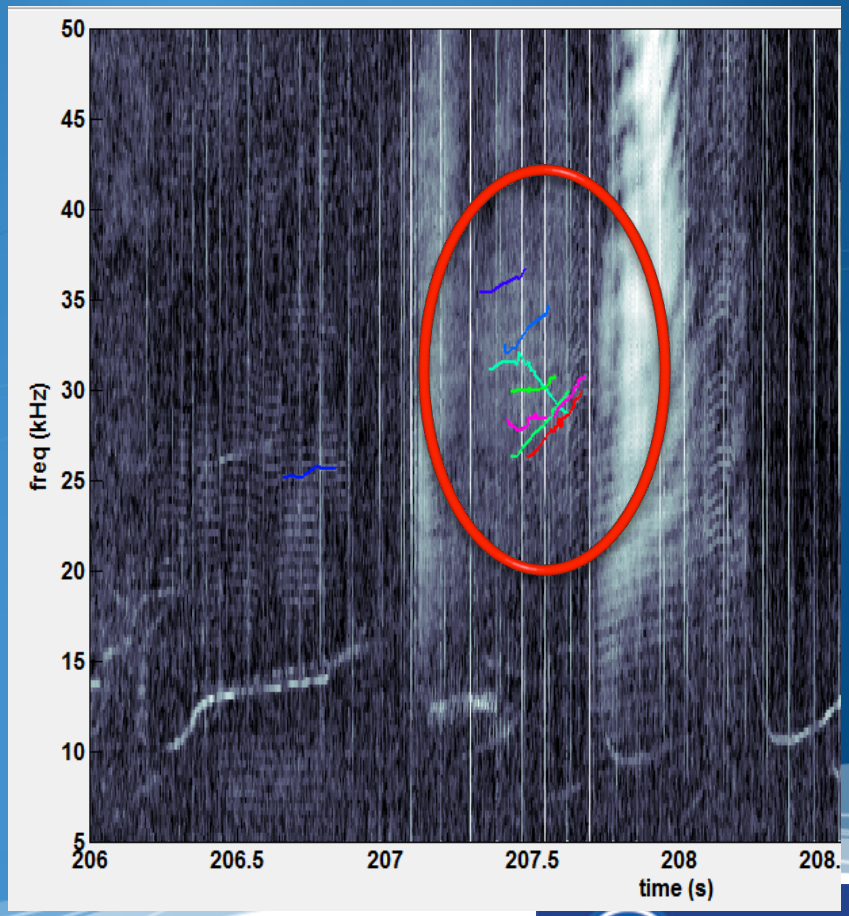


Bio·Waves Incorporated

Presence of clicks and buzzes

- Increased number of missed whistles, false positives





Bio·Waves Incorporated

What does it all mean?

- All detectors performed well and had different strengths and weaknesses
  - PAMGuard: highest precision and recall, ☺
    highest fragmentation and frequency deviation ☹
  - *Silbido*: lowest fragmentation, highest % coverage ☺
    - New version of *Silbido* – higher precision and recall
  - Ishmael: lowest frequency deviation, ☺
    lowest recall ☹

Bio·Waves Incorporated

What does it all mean?

- Choice of detector depends on analysis goals
  - Detection only?
    - PAMGuard
  - Detection and classification?
    - *Silbido*
- Important to understand detector performance
  - Groundtruth on all new datasets

Bio·Waves Incorporated

Acknowledgements

- Tina Yack, Elizabeth Ferguson, Tom Norris, Bio-Waves, Inc.
- Yvonne Barkley and Roch lab
- Mike Weise, ONR
- Anurag Kumar and Bob Gisiner, LMR

Bio·Waves Incorporated


Thank you!