

Data Cleansing with Contrastive Learning for Vocal Note Event Annotations

Gabriel Meseguer-Brocal¹, Rachel Bittner², Simon Durand², Brian Brost²

1- Ircam Lab, CNRS, Sorbonne Université Paris, France, Paris.

2- Spotify, USA.

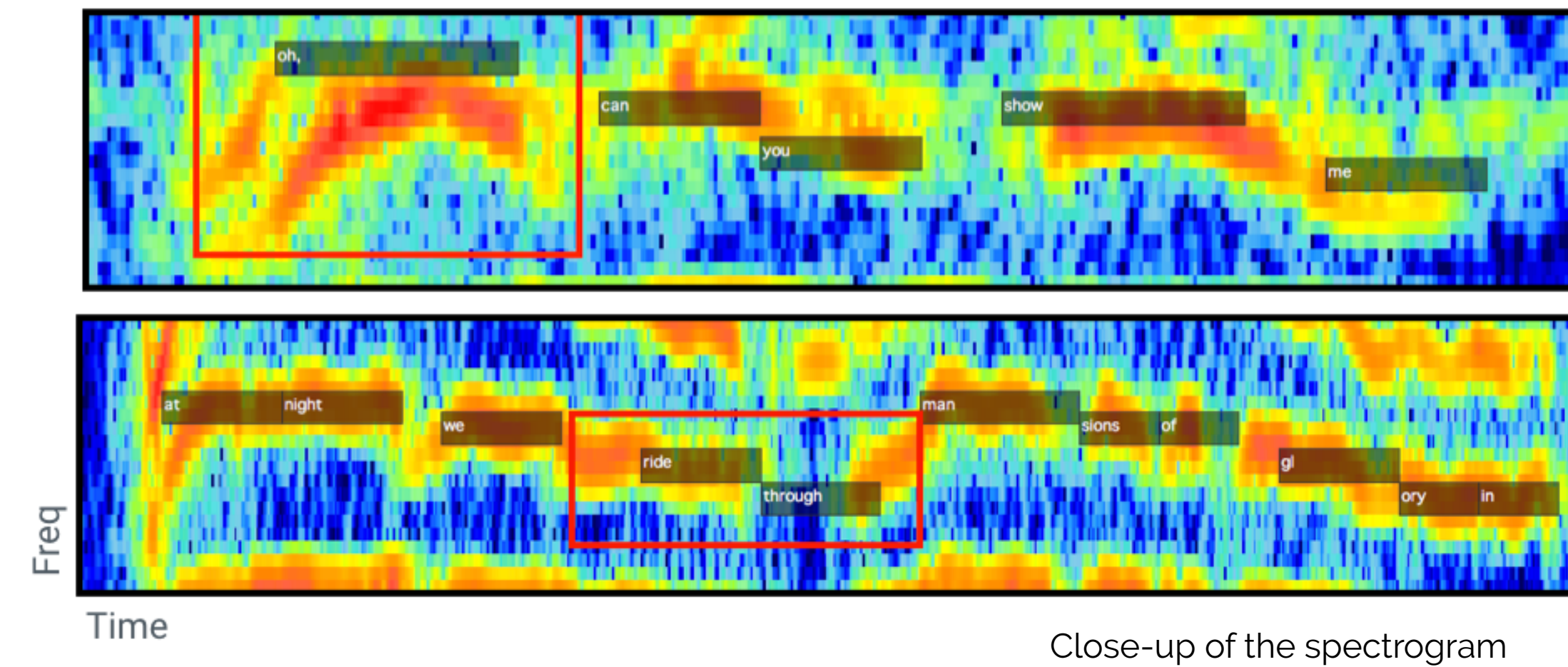
gabriel.meseguerbrocal@ircam.fr, rachelbittner@spotify.com, durand@spotify.com, brianbrost@spotify.com

Music data

Most MIR problems are solved using supervised methods → labeled data.

Music datasets have label errors:

- Incorrect labels.
- Imprecise event times.
- Missing/extra events.



Particularly problematic for datasets derived from the Internet.

Bad annotations are problematic for *training* and *evaluation*.

What can we do about it?

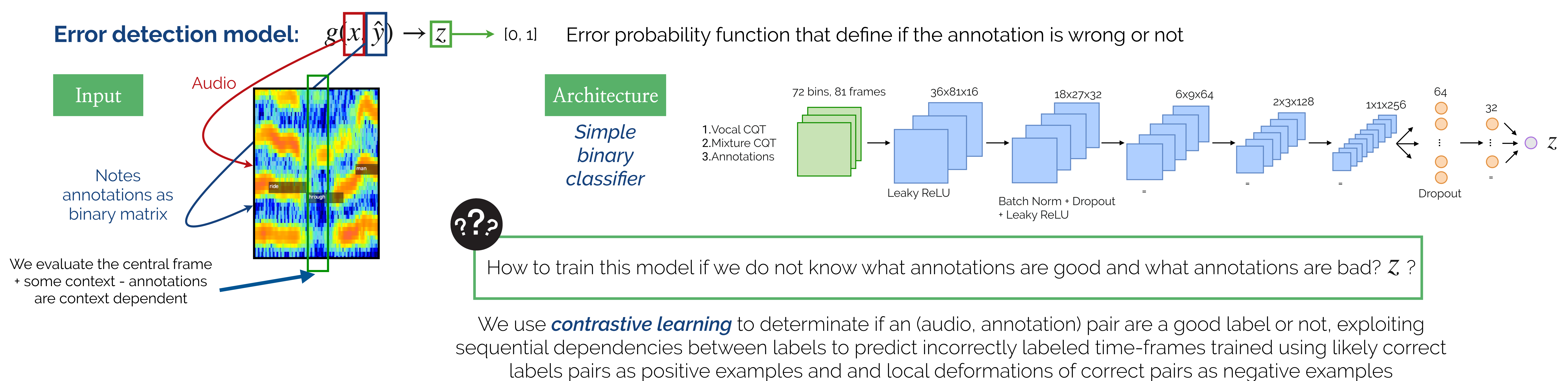
- Getting more *human* annotations: demanding and costly.
- Automatically *correcting* the errors: difficult for complex tasks.
- Find the errors in an automatic way.

Data cleansing studies how to mitigate the effects of label noise.

Case study

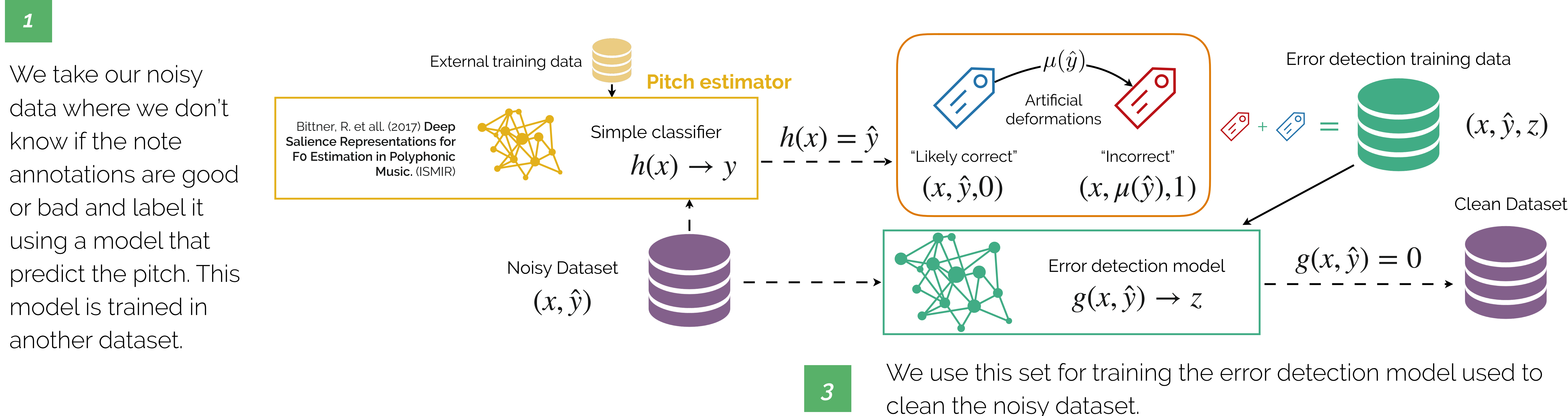
Note annotations → errors in time and frequency.

Our solution

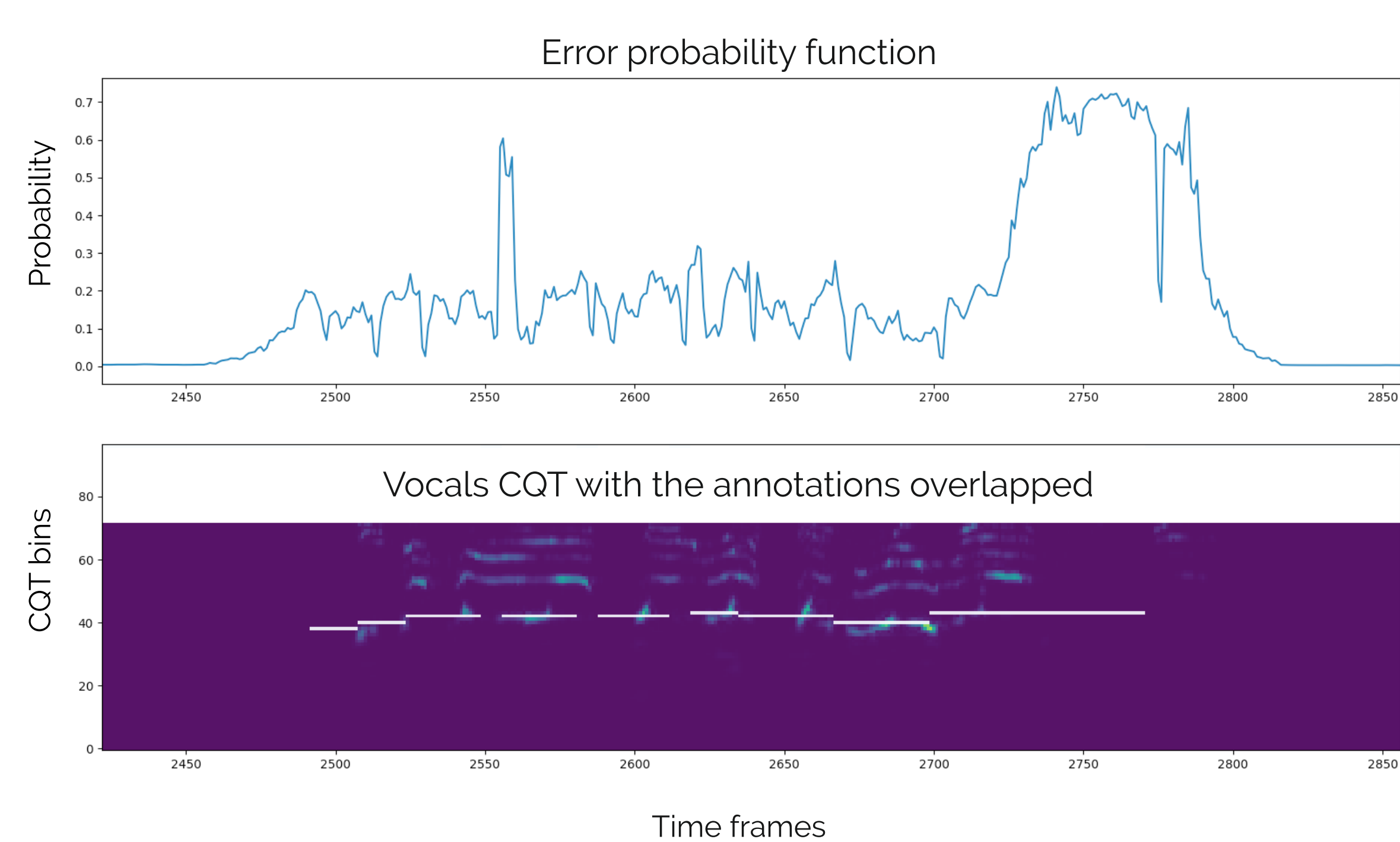


Contrastive learning

- 1
- 2 The pitch labels are close to the notes annotation we have in the noise dataet. By comparing both, pitch estimations and note annotations, we can select the "likely correct" examples, where the prediction is similar to the label. We distort them to generate the incorrect examples, defining our training set.

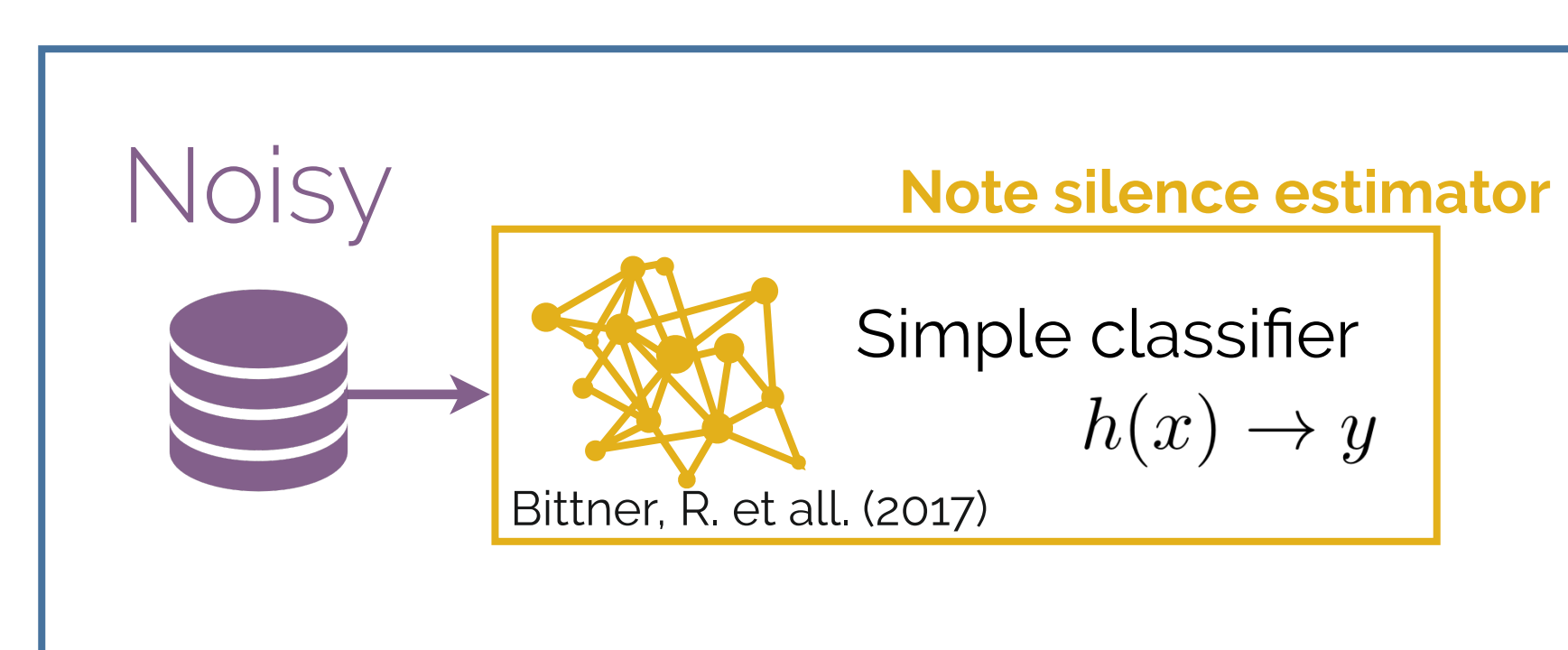


Results

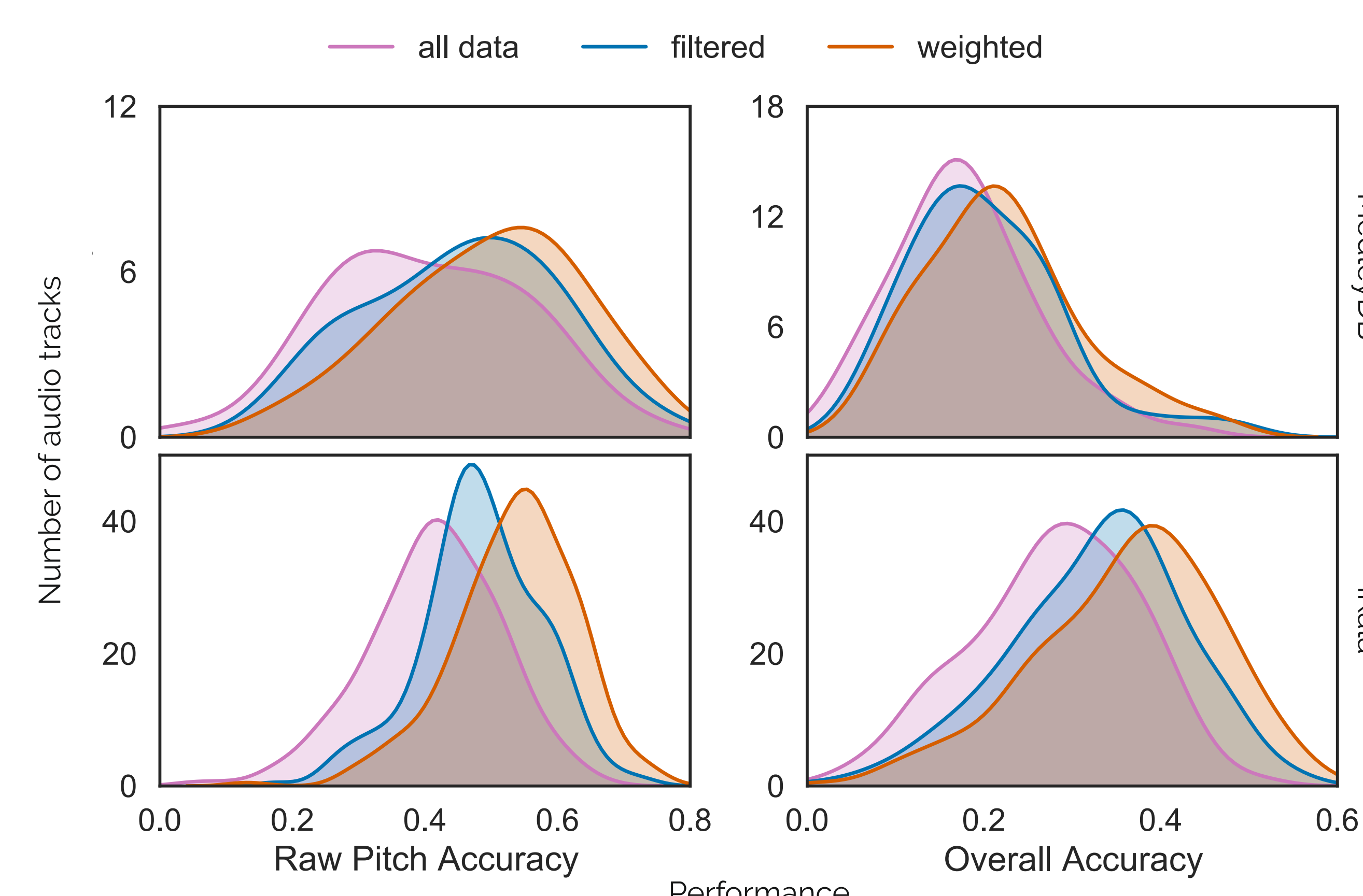


How to validate $g(x, \hat{y}) \rightarrow z$?

- a. Directly: no "real" ground truth (only likely correct).
- b. Manually: costly and required expert knowledge.
- c. **Data cleansing:** identifying incorrect annotations and remove them during training.



- x3 - different training conditions
1. All
 2. Filtered $z \leq 0.5$
 3. Weighted loss by $1 - z$



Distribution of scores for the three training conditions. The curves are shifted to the right which means we have better results for the models trained after filtering the bad annotations with our error probability function.