

# ECS7013: Deep Learning for Audio and Music

## Lab 5 – Event Annotation

In this lab, we will experiment with audio event annotation which is an important step to train a frame-based “event detection” or “transcription” system.

### 1. Annotation to frame-wise labels

More specifically, you will transform some annotation data into a format (for example, spectrogram with frame-wise labels in Figure 1) useful for training. Note that you will work with one-hot encoding labels rather than discrete labels shown in Figure 1.

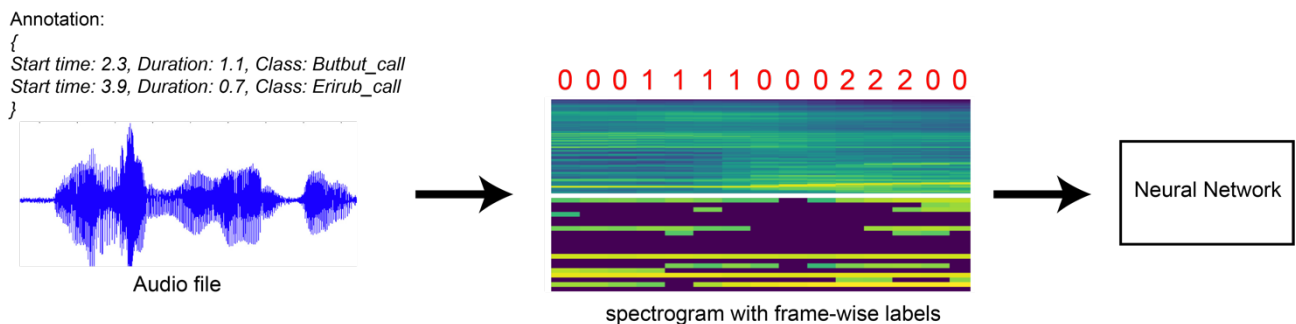


Figure 1 Transform event annotation into frame-wise labels.

### 2. Frame-wise labels to annotation

In testing, presented with a unseen spectrogram, the trained network predicts and assigns labels to frames of the spectrogram. You will then need to transform the predicted frame-wise labels into the human annotation format. This procedure is illustrated in Figure 2. Note that you will work with one-hot encoding frame-wise labels rather than discrete ones shown in the figure.

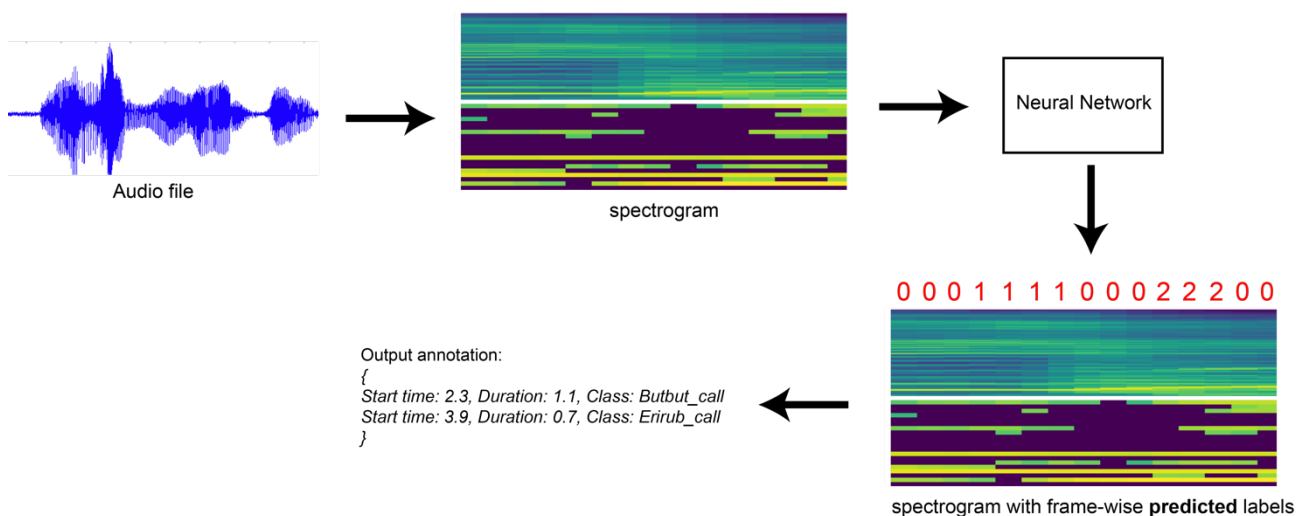


Figure 2 Using the trained network for evaluation and transform the predicted frame-wise labels to human-readable annotation.