

# Speech Quality in Media

Jayson DeLancey  
Dolby Laboratories  
San Francisco, USA  
jayson.delancey@dolby.com

Joan Serrà  
Dolby Laboratories  
Barcelona, Spain  
joan.serra@dolby.com

## ABSTRACT

In this workshop, we describe building a project integrating Web Audio with REST APIs. Highlighted will be discussion of an approach for quantifying audio quality. This semi-supervised algorithm helps assess changes in speech-based audio quality.

## 1. WORKSHOP

### 1.1 Overview

Automatic speech quality assessment has many real-world applications. To help support some of these use cases Dolby has built a developer platform to make these algorithms available as REST APIs.

### 1.2 Algorithm

In this part we will detail a bit the mechanism behind speech quality assessment, which is largely powered by our recent algorithm SESQA [1]. SESQA is a reference-free semi-supervised model based on neural networks. SESQA can assess multiple aspects of speech quality simultaneously from single-audio utterances (as opposed to reference-based systems, which need an additional clean reference), and it does so by learning from both labeled and unlabeled data (as opposed to learning just from labeled/annotated data). Labeled data comes from crowd-sourced human judgments of quality and unlabeled data corresponds to programmatically-generated distortions that are used at training-time to teach the network about ranking and consistency relations, two notions that are found to complement correlation with human judgment.

### 1.3 Media APIs

Use cases for the algorithms discussed generally fall into file-based media processing and real-time interactive communications.

Postman is a web-based API development tool to build, test, and modify APIs including making HTTP requests. During the workshop we'll review collections in Postman that demonstrate some methods of exploring and visualizing data when prototyping new applications by combining

REST APIs for more data-intensive algorithms and integrate with Web Audio APIs for visualization and exploration of that media.

## 2. SPEAKER BIOS

**Jayson DeLancey** — Jayson DeLancey manages Developer Relations for Dolby.io in San Francisco (since 2019). He has 15+ years of experience as a software engineer with companies including The MathWorks (MATLAB), DreamWorks Animation, and General Electric. His area of focus is to help enable software developers and creative media professionals to build new applications.

**Joan Serrà** — Joan Serrà is a Staff Researcher with Dolby Labs in Barcelona (since 2019), where he works on deep learning for audio processing and understanding. He did an MSc (2007) and PhD (2011) in machine learning for audio at the Music Technology Group of Universitat Pompeu Fabra, and a postdoc in artificial intelligence at IIIA-CSIC (2011–2015). After that, he was a machine learning researcher at Telefónica R&D (2015–2019). Joan has been involved in several research projects, funded by National and European institutions, and co-authored over 100 scientific publications, many of them highly-cited and in top-tier journals and conferences. He occasionally acts as reviewer for some of those and other venues and often gives talks and lectures on a variety of deep learning topics.

## 3. REFERENCES

- [1] J. Serrà, J. Pons, and S. Pascual. SESQA: Semi-supervised learning for speech quality assessment. In *Proc. of the IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, pages 381–385, 2021.



Licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0). **Attribution:** owner/author(s).

*Web Audio Conference WAC-2021, July 5–7, 2021, Barcelona, Spain.*

© 2021 Copyright held by the owner/author(s).