

# Binaural Scene Classification with Wavelet Scattering

Vincent Lostanlen and Joakim Andén

École normale supérieure  
45 rue d'Ulm, 75005 Paris, France

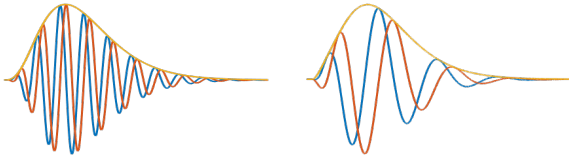


Fig. 1. Left: Gammatone wavelet  $\psi_{\gamma_1}(t)$  with a quality factor of  $Q_1 = 8$ . Right: Gammatone wavelet  $\psi_{\gamma_2}(t)$  with a quality factor of  $Q_2 = 1$ . Blue and red oscillations represent the real and imaginary parts. The orange envelope represents the complex modulus.

**Abstract**—This technical report describes our participation to the scene classification task of the 2016 edition of the IEEE AASP Challenge for Detection and Classification of Acoustic Scenes and Events (DCASE). Our computational pipeline consists of a Gammatone scattering transform, local averaging to a time scale of 740 ms, logarithmic compression, and frame-based classification with a linear support vector machine (SVM). At test time, decisions are aggregated over the duration of the full auditory scene by majority voting. We propose a novel data augmentation technique, by mixing the left and right stereophonic channels in varying proportions, in order to enforce invariance to azimuthal orientation of the binaural recording system.

## I. INTRODUCTION

Invariance to translation is of paramount importance in auditory scene analysis. Indeed, the origin of time is chosen arbitrarily by the recordist, and thus do not convey any information about the class.

Scattering transforms are translation-invariant representations of audio signals which cascade an auditory filterbank and a modulation filterbank, interspersed with complex modulus nonlinearities.

## II. SCATTERING TRANSFORM

## III. DATA AUGMENTATION

## IV. DISCUSSION

## V. CONCLUSION

This technical report presents

TABLE I  
CROSS-VALIDATION RESULTS.

Scene	Baseline	Temporal scattering (1 azimuth)	Temporal scattering (5 azimuths)
beach	72.0 ± 19.0		84.9 ± 11.7
bus	62.0 ± 19.6		91.3 ± 14.4
café/restaurant	83.9 ± 12.1		57.1 ± 20.5
car	75.7 ± 17.7		93.7 ± 6.3
city center	85.6 ± 18.5		92.8 ± 8.9
forest path	65.9 ± 26.5		96.4 ± 7.2
grocery store	76.6 ± 22.3		84.3 ± 15.6
home	79.5 ± 18.6		64.2 ± 26.3
library	61.3 ± 30.5		81.1 ± 7.5
metro station	85.3 ± 18.0		98.7 ± 2.7
office	91.1 ± 7.9		86.1 ± 27.8
park	24.5 ± 15.6		75.6 ± 10.5
residential area	75.4 ± 29.1		55.8 ± 20.2
train	36.7 ± 10.8		57.1 ± 15.3
tram	89.5 ± 17.5		86.4 ± 11.7
average	71.3 ± 3.8		80.4 ± 3.5