[7]

In Adavanne et al, LSTM units (Long short-term memory) within a Recurrent Neural Network (RNN) are leveraged to identify sound events in polyphonic audio samples [7]. Three different feature sets were studied: log mel-band energy, harmonic features (such as pitch), and time difference of arrival (TDOA). A combination of mel-band energy and TDOA features performed the best in this context.

[8] and [9]

Royo-Letelier et al [8] explored the use of metric learning to disambiguate artists within a music catalog. They found that at smaller dataset sizes (300 or fewer examples), that a traditional 1D-CNN (One Dimensional Convolutional Neural Network) outperformed the metric learning model [8]. The model they compared their metric learning model to is described in [9].

In [8], triplet loss was used to separate different groups of musicians. This loss mechanism learns a metric preserving map *f* that groups similar pairs and distances dissimilar pairs given a triplet of (xa, x+, x-).

Volker B. Deecke and Vincent M. Janik. “Automated categorization of bioacoustics signals: Avoiding perceptual pitfalls.” Journal of the Acoustical Society of America, vol. 119, no. 1, pp. 645-653, 2006.

The authors present data and results of using unsupervised learning to categorize animal vocalizations into meaningful biologically relevant categories. Used ‘known’ data from bottlenose dolphins and orca whales. The data was categorized by and agreed upon by humans specialists. Algorithm only referenced and not provided in paper. Used dynamic time warping and adaptive resonance theory neural networks of ‘frequency contours’ of the biological acoustic signals. These ideas leverage the natural features of sound by vertebrate auditory perception:

1. Relatively insensitive to changes in the duration of a vocalization pattern. Animals don’t care if the vocalization duration is different by small amounts, more sensitive to frequency changes.
2. Frequency perception is nonlinear. Frequency perception is on logarithmic scale, octaves.

Need to allow for adjusting the duration of a signal feature when comparing to a known example, and work with octave frequencies. The adaptive resonance theory 2 neural network compares a signal to a set of references, if close categorize with reference, else make new reference category. Demonstrated improved accuracy of categorization as compared to human observers and improved consistency. The automated method was returned the same categorization for a given data set.

* Supports the use of octave filters

**Wavelets**

Historically, much research into audio signal classification using joint time-frequency representation techniques has involved the traditional Fourier Transform or time-dependent variants such as the Short-Term Fourier Transform (STFT). While such analysis via the Fourier basis offers decent feature extraction results, there exist alternative methods that may offer superior results -- primarily with respect to the acoustic domain and human perception signals where frequencies of interest can lie an octave, dyadic scale.

Here we consider the use of wavelets and the discrete wavelet and wavelet packet transforms as techniques for audio signal feature extraction and classification. A key difference between wavelet transforms and Fourier transforms is that while Fourier transforms have a single set of basis functions in sine, cosine, the wavelet transforms do not; they instead have an infinite number of basis functions stemming from an appropriately chosen “mother wavelet”, and scaling and shifting of that wavelet function. The result of this is that the 2-D time-frequency joint space can be decimated into a multiresolution analysis, enabling detailed observations at certain times *and* frequencies otherwise obscured from a “uniform” resolution analysis, such as that from the STFT. [https://www.di.ens.fr/~mallat/papiers/MallatTheory89.pdf] [Graps]. Furthermore the localization and finite energy properties) of wavelets allows the DWT family to be especially suited for signals of time-varying frequency structure signals such as audio recordings (\*\*\*)

Insofar the DWT has been primarily used in applications such as signal denoising, compression, and image feature detection. It has also been widely exploited in biomedical signal processing, such as in classification of transient content of EEG signals for neurological purposes. The literature for DWT-based audio signal feature extraction and classification is relatively minimal; the following feature sets for environmental noise contimation of aircraft acoustic recordings were …………..<**unfinished and unedited. Need to make more succinct> .**

In [ ], D. Percival formalizes the “wavelet variance” – a decomposition of the variance of a time-series process across different “scales.” The work shows how the wavelet variance of a time series signal can be decomposed into individual functions of scale (i.e. time) analogously to how a real-valued stochastic process’s variance – the spectrum – decomposes across frequencies. In this regard the wavelet variance at each scale is a measure of “measure of how much a weighted average with band-width λ of the process {Yt} changes from one time period of length λ to the next” []. The wavelet variance is furthermore a compact octave-band representation of the signal, capturing the time-varying frequency charatistics per octave into a single value per passband, as under most conditions, the wavelet variance writes as for the stationary time-varying process/signal *S*.

LI and Tao propose entropy-based features stemming from the discrete wavelet packet transform (DWPT). They suggest that given choice of mother wavelet and decomposition level *L* to form a feature vector by concatenating all Shannon entropies of the terminal nodes of the DWPT [], where the Shannon entropy here is a measure of uncertainty of a random variable in information theory based on the probability distribution of wavelet packet energy at the nodes per level of the transform. The DWPT is similar to the DWT: both can be viewed as a multirate filter bank of applying pyramidal structure of low and high pass filters (convolving with scaled/shifted versions of the mother wavelet) to the signal to achieve a binary tree of height *L*. However while the DWT cascades filters on only the approximation coefficients (low-pass output) and thereby produces the dyadic decomposition, the DWPT cascades on both outputs producing a balanced binary tree with 2L nodes/coefficients. [CITATION NEEDED]

Tzanetakis, Essl, et. Al discuss methods to classify non-speech audio signals and beat detection using the Discrete Wavelet Transform. In particular they highlight features generated from statistics about the coefficients in each dyadic sub-band. These are for a band level *j*: the mean of the absolute values of *jth*-level coefficients, standard deviation of *jth*-level coefficients, and vector of ratios between means between the *jth* and *jth + 1* subbands for *j = 1, 2, …, L.*

[1] Tzanetakis, George & Essl, Georg & Cook, Perry. (2001). Audio Analysis using the Discrete Wavelet Transform. Proceedings of the Conference in Acoustics and Music Theory Applications. 318-323.

[2] Li T, Zhou M. ECG Classification Using Wavelet Packet Entropy and Random Forests. Entropy. 2016; 18(8):285.

[3] Percival, Donald. (2012). On Estimation of the Wavelet Variance. Biometrika. 82. 10.1093/biomet/82.3.619.

[4] Graps, Amara. (1995). "An Introduction to Wavelets". IEEE Comp. Sci. Engi.. 2. 50-61. 10.1109/99.388960.

[5] Hazarika, Neep & Z. Chen, J & Chung Tsoi, Ah & Sergejew, Alex. (1997). Classification of EEG signals using the wavelet transform. Signal Processing. 59. 89-92 vol.1. 10.1109/ICDSP.1997.627975.

[6] Mallat paper…

# References

Code of Federal Regulations. (2016). Title 14, Part 36, Noise Standards: Aircraft Type and Airworthiness Certification. Washington, D.C.: Federal Aviation Adminstration.

International Civil Aviation Organization. (2014). Standards and Recommended Practices, Annex 16. *Vol. 1, Aircraft Noise, 7th*. Montreal, Canada: International Civil Aviation Organization.

From before

# References

|  |  |
| --- | --- |
| [1] | Federal Aviation Administration, "Aircraft Noise Issues," [Online]. Available: https://www.faa.gov/about/office\_org/headquarters\_offices/apl/noise\_emissions/airport\_aircraft\_noise\_issues/. [Accessed 1 12 2018]. |
| [2] | US National Parks Service, "Sounds Gallery," [Online]. Available: https://www.nps.gov/subjects/sound/gallery.htm. [Accessed 2 12 2018]. |
| [3] | Y. Wang, X. Wu, X. Li and J. Zhou, "Vehicle type recognition in sensor networks using improved time encoding signal processing algorithm," *Mathematical Problems in Engineering,* vol. 2014, 2014. |
| [4] | H. Wu, M. Siegel and P. Khosla, "Vehicle sound signature recognition by frequency vector principal component analysis," in *Instrumentation and Measurement Technology Conference*, 1998. |
| [5] | J. Andén and S. Mallat, "Deep scattering spectrum," *IEEE Transactions on SIgnal Processing,* vol. 62, no. 16, pp. 4114-4128, 2014. |
| [6] | F. Hemmati, W. Orfali and M. Gadala, "Roller bearing acoustic signature extraction by wavelet packet transform, applications in fault detection and size estimation," *Applied Acoustics,* vol. 104, pp. 101-118, 2016. |
| [7] | S. Adavanne, G. Parascandolo, P. Pertilä, T. Heittola and T. Virtanen, "Sound event detection in multichannel audio using spatial and harmonic features," in *Detection and Classification of Acoustic Scenes and Events*, Budapest, Hungary, 2016. |
| [8] | J. Royo-Letelier, R. Hennequin, V. Tran and M. Moussallam, "Disambiguating music artists at scale with audio metric learning," in *19th International Society for Music Information Retrieval Conference*, Paris, 2018. |
| [9] | J. Park, J. Lee, J. Park, J. Ha and J. Nam, "Representation learning of music using artist labels," in *19th International Society for Music Information Retrieval Conference*, Pairs, 2018. |
| [10] | Atlassian, "Agile Tools," [Online]. Available: https://trello.com/power-ups/59d4ef8cfea15a55b0086614/agile-tools. [Accessed 1 12 2018]. |