[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**

*Still very much scratch work.*

It is well-known that Fourier and other spectral analyses are cornerstone to signal processing. The Fourier Transform, for example, decomposes a signal into a set of sine and cosine basis functions of different frequencies, thereby transforming a given signal into a corresponding frequency domain representation. Feature extraction can thusly be deployed by analyzing key characteristics – such as ‘spectral’ energy in various passbands or FFT coefficients at important frequencies.

However, signal classification techniques using feature extraction via the Fourier Transform pose one limitation: the time-axis is ignored. Namely, using solely Fourier techniques on the entire signal prevent a full comprehensive analysis of signals such as audio, where time is a crucial factor – we seek to examine “where” and “when” important spectral coefficients emerge. This notion is the uncertainity principle in signal processing: the cost of higher resolution in time domain is lower resolution in the frequency domain, and vice versa

A solution to the joint time-frequency analysis problem has been the Short-Term Fourier Transform (STFT), which injects a sliding window function into the main equation of the Discrete Fourier Transform. However, the STFT resultantly creates uniformity in the time vs. frequency “plane.” This can be disadvantageous in certain applications, especially audio signal processing and human acoustics, where frequencies of interest lie on an octave/dyadic “logarithmic” scale.

A more modern approach to the joint time-frequency problem has been the Wavelet Transform.

The Discrete Wavelet Transform (DWT) and Discrete Wavelet Packet Transform (DWPT) have been

The following features are used in our system:

* The mean of the absolute value of the coefficients in each subband. These features provide information about the frequency  distribution of the audio signal.
* The standard deviation of the coefficients in

each subband. These features provide information about the amount of change of the frequency distribution

ï Ratios of the mean values between adjacent subbands. These features also provide information about the frequency distribution. (averaging approach across the possible resolutions)

approximation and lowest detail level, so I suspected that this was overkill. To test this hypothesis, I ran classification experiments using the Daubechies wavelet family at various decomposition levels for feature extraction. The resulting success rate at each of 8 different levels, averaged across db1 through db8 wavelets, is shown in Figure 5. Unlike the MFCCs, where the success rate was not a smooth function of the number of coefficients, with the wavelet decomposition there is a reasonably smooth increase in success rate as the number decomposition levels increases, up until the highest number of levels, where as expected, the small size of the resulting final band was likely more misleading than helpful. While the highest average success rate occurred at 11 levels, the highest rate for an individual wavelet was the db6 wavelet at 8 levels (71.33%). The db4 wavelet at 8 levels was almost as effective as db6, with 68.67% success. The highest success rate at 11 levels was 69.33%, so even though the average success rate was higher elsewhere, I chose the db4 wavelet with an 8-level decomposition as the ideal balance of computational complexity (filter length and decomposition levels) and success rate. Later comparisons with MFCCs were all performed using this wavelet.

(Michelle)

Matlab – specific decomposition level – spectral energies and

**References to Add:**

* Audio Analysis using the Discrete Wavelet Transform. George Tzanetakis, Georg Essl, Perry Cook
* Classification of Percussive Sounds Using Wavelet-Based Features Michelle Daniels
* Feature Extraction and Reduction of Wavelet Transform Coefficients for EMG Pattern Classification. A. Phinyomark, A. Nuidod, P. Phukpattaranont, C. Limsakul
* <https://www.mathworks.com/examples/wavelet/mw/wavelet-ex52408711-signal-classification-using-wavelet-based-features-and-support-vector-machines>
* <https://www.eecis.udel.edu/~amer/CISC651/IEEEwavelet.pdf>

# 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]. |