NON-STATIONARY SIGNAL CLASSIFICATION USING JOINT FREQUENCY ANALYSIS

Somsak Sukittanon [1], Les E. Atlas [1], James W. Pitton [2], and Jack McLaughlin [2]

[1] Department of Electrical Engineering, [2] Applied Physics Laboratory (APL)

University of Washington, Seattle, WA

ABSTRACT

Time-varying short-term spectral estimates have been successfully applied in many classification tasks. However, they are still insufficient for many non-stationary signals where time-varying information is useful. In this paper, we propose to improve the deficiencies of current short-term feature analysis by adding information to describe the time-varying behavior of the signals. Our proposed method which is motivated by the human auditory system can be applied to several non-stationary signal types. Real world communication signals were used for experimental verification. These experimental results, assessed with a conventional probabilistic classifier, showed significant improvement when the new features were added to short-term spectral estimates.

1. INTRODUCTION

In automatic classification, the goal is to have a machine characterize events and make an appropriate decision about the class of these events. Short-term estimates, for example Fourier or autocorrelation coefficients, have been widely used for many signal types (e.g. [1, 2]). Basically, the signal is blocked into successive frames using a small data analysis window. This blocking assumes stationarity of the signal within each frame. The windowed signal is then transformed into a new representation hopefully giving good discrimination and/or energy compaction. The length of a short-term analysis window can be different depending on the signal type. For example in tool-wear classification [3], stationarity can be assumed within a quarter revolution of a tool. Since a working tool has different sizes and rotation speeds, a quarter revolution time window results in different time durations, 20-40 ms. In speech recognition [4], the typical short-term window used for speech analysis is about 20-30 ms. In music classification, the length of a short-term window is influenced by audio coding where typically two processing windows are typically used [5]. With a sampling rate of 48 KHz, windows of size 256 samples, or about 5 ms, and 2048 samples, or about 43 ms, are commonly applied. Unless specified otherwise, we will refer a data window of length less than 50 ms as a short-term window.

In this paper, we model non-stationary signals, such as speech, music, or communication signals, as the product of a narrow bandwidth lowpass process $m(t)$ modulating a higher bandwidth carrier $c(t)$, as shown in Equation (1). To effectively use this model, $m(t)$ is assumed to be nonnegative and its bandwidth does not overlap with $c(t)$.

$$x(t) = m(t)c(t) \quad (1)$$

The above model has been applied to encode $x(t)$ in speech [6] and audio [7]. Motivated from the success in coding, this model can also be useful for non-stationary signal classification. An important question can be raised whether more traditional feature extraction such as short-term spectral analysis is adequate for extracting the pertinent information of this model. Since $m(t)$ is a slowly varying signal, using too short of an analysis window can be insufficient to model $m(t)$.

Understanding the human recognition system and integrating the relevant aspects significantly contributes to the understanding of this form of feature extraction. For example in [4], to estimate energy at the most sensitive modulation frequency of human audition, about 4 Hz, an analysis window of at least 250 ms is needed. A related concept on a perceptual duration of auditory system is a pre-perceptual auditory image [8]. This concept refers to the process where an auditory input produces an auditory image containing information about a stimulus. Because of the continuous change of the auditory input, a pre-perceptual auditory store is used to hold information of the stimulus and can be utilized later. Massaro [8] proposed the estimation of this perceptual unit using a backward masking experiment. Results suggested that pre-perceptual auditory storage and processing was over 200 ms which is again longer than a short-term analysis window. Finally, the sensitivity of short-term features to noise [9] and unseen testing data [3] is another deficiency.

To improve the deficiencies of short-term feature analysis, we propose long-term feature analysis called joint frequency analysis. Joint frequency analysis not only contains short-term information about the signal, but also contains long-term information representing patterns of time variation. We will show that our proposed feature analysis is complimentary to more traditional short-term feature analysis. Communication signal classification is used for experimental verification. Related work [10] using joint frequency analysis in a communication signal interception application used a method requiring *a priori* information such as a symbol rate. To avoid the assumption of prior information, we use time-frequency theory integrated with psychoacoustic results on modulation frequency perception to provide a foundation and subsequent joint frequency representation for a classification system.

2. JOINT ACOUSTIC AND MODULATION FREQUENCY ANALYSIS

2.1 Theory

One possible joint frequency representation, $P(\eta,\omega)$, is a transform in time of a demodulated short-time spectral estimate. For the purpose of this paper, ω and η are "acoustic frequency" and "modulation frequency," respectively. A spectrogram, or other joint time-frequency representations [11], can be used as the starting point of this analysis. In this paper, we first use a spectrogram with an appropriately chosen window length to estimate a joint time-frequency representation of the signal, $P_x^{SP}(t,\omega)$. Second, another transform (e.g. Fourier) is applied along the time dimension of the spectrogram to estimate $P_x^{SP}(\eta,\omega)$. Another way of viewing $P_x^{SP}(\eta,\omega)$, as shown in Equation (2), is the convolution in ω and multiplication in η of the correlation function of a Fourier transform of the signal $x(t)$ and the underlying data analysis window $h(t)$:

$$P_x^{SP}(\eta,\omega) \;=\; \left(H^*(\omega-\frac{\eta}{2})H(\omega+\frac{\eta}{2})\right) *_\omega \left(X^*(\omega-\frac{\eta}{2})X(\omega+\frac{\eta}{2})\right). \qquad (2)$$

To illustrate the behavior of $P_x^{SP}(\eta,\omega)$, an AM signal is used. Equation (3) describes a spectrogram of the AM signal, where ω_m and ω_c are the modulation and carrier frequencies, respectively.

$$P_x^{SP}(t,\omega) \;\simeq\; \left(1+\cos\omega_m t\right)^2 \delta(\omega\pm\omega_c) \qquad (3)$$

Applying a Fourier transform along the time dimension of $P_x^{SP}(t,\omega)$:

$$\begin{aligned} P_x^{SP}(\eta,\omega) &= \int_{-\infty}^{\infty} P_x^{SP}(t,\omega)e^{-j\eta t}dt \\ &\simeq 2\pi\left(\frac{3}{2}\delta(\eta)+\delta(\eta\pm\omega_m)+\frac{1}{4}\delta(\eta\pm 2\omega_m)\right)\delta(\omega\pm\omega_c). \end{aligned} \qquad (4)$$

$P_x^{SP}(\eta,\omega)$ results in the compaction of non-zero values occurring at low η, $|\eta|\le 2\omega_m$, which can be advantageous for coding and classification. When using a joint frequency representation as features in non-stationary signal classification, only positive and low modulation frequencies are needed for estimating the low-dimensional features because of the symmetric property in frequency and small non-zero support region of $P_x^{SP}(\eta,\omega)$. The modulation frequency range of interest in $P_x^{SP}(\eta,\omega)$ can be determined by the assumed highest modulation frequency in the signal or the bandwidth of the chosen spectrogram window.

2.2 Interpretation

In Equation (2), $P_x^{SP}(\eta,\omega)$ values along $\eta=0$ are the convolution of the spectra of the spectrogram window and the signal.

$$P_x^{SP}(0,\omega) \;=\; |H(\omega)|^2 *_\omega |X(\omega)|^2 \qquad (5)$$

The values lying along $\eta=0$ are an averaged short-term spectral estimate of the signal. The length of the spectrogram window and the amount of overlap between data analysis windows determine the trade-offs between the bias and variance of the short-term spectral estimate. $P_x^{SP}(\eta,\omega)$ at $\eta=0$ is an estimate of the stationary information while the $P_x^{SP}(\eta,\omega)$ at $\eta\neq 0$ is an estimate of the non-stationary information about the signal. This non-stationary information can represent various quantities (e.g. a symbol rate of a communication signal).

In this paper, joint frequency analysis is applied to the automatic classification of unknown communication signals. For digital communication signals such as frequency shift keying (FSK) or phase shift keying (PSK), the message is encoded as the change of frequency or phase of the carrier, respectively. For the spectrogram of FSK as shown in Figure 1a, the transmitted signal is sent by switching between two frequencies, 1200 Hz and 1500 Hz. Since filterbank analysis is effected via a spectrogram, the change of frequency is transformed into a change of magnitude in each subband of the joint time-frequency representation. The non-zero terms in $P_x^{SP}(\eta,\omega)$ as shown in Figure 1b reveal how fast the instantaneous frequency of this signal pass through the subband filters. The non-zero terms occurring at harmonics of approximately 30 Hz in the η dimension of $P_x^{SP}(\eta,\omega)$ reflect the symbol rate of this FSK signal. For the PSK signal, the phase change of the carrier also contributes to the change of the instantaneous amplitude. The spectrogram of PSK as illustrated in Figure 2a shows the signal having high bandwidth in ω and random behavior in time. $P_x^{SP}(\eta,\omega)$ of this PSK signal, as shown in Figure 2b, exhibits a more compact energy representation. There are non-zero terms energy representation. There are non-zero terms occurring at harmonics of approximately 15 Hz in the η dimension. For a multilevel modulation signal such as multilevel FSK (MFSK), the non-zero terms will appear in more acoustic frequency subbands than an FSK signal.

As demonstrated with real-world signals, joint frequency analysis has the potential to extract time-varying information via the non-zero terms in the representation. These non-zero terms are possibly useful for discriminating signal types, thus they should be considered as useful features. However, using $P_x^{SP}(\eta,\omega)$ directly for classification has an important disadvantage. $P_x^{SP}(\eta,\omega)$, as well as other two-dimensional analysis, is a long-term analysis, therefore it provides an extremely large dimension compared to traditional short-term spectral estimate. Even though we can reduce the feature dimension due to the symmetry in frequency and small non-zero support region of $P_x^{SP}(\eta,\omega)$, the resulting dimensions are still too large for typical classifiers. Past research has addressed the method of reducing feature dimension of a two dimensional representation in various ways. For example, many methods view a two-dimensional signal representation as an image. The non-zero terms lying in the representation then can be viewed as the lines or objects and the small set of descriptors being invariant to translation, rotation, or scaling can be extracted. Since we are interested in tasks where human auditory signal classification is largely successful, integrating psychoacoustic results into the analysis can possibly provide added advantages in feature design and selection.

2.3 Modulation scale

Using Fourier analysis for the modulation frequency transform in the above analysis results in a uniform frequency bandwidth in modulation frequency; however this approach for modulation decomposition can be inefficient for auditory classification due to the resulting high dimensionality. Furthermore, the uniform bandwidth in modulation frequency does not mimic the human auditory system. Recent psychoacoustic results [12] suggest that a log frequency scale, with a constant-Q over the entire frequency range, best mimics human perception of modulation frequency. Our approach uses a continuous wavelet transform (CWT) to efficiently approximate this constant-Q effect. Our method, called modulation scale analysis, starts with a standard spectrogram of $x(t)$:

$$P_x^{SP}(t,\omega) \;=\; \frac{1}{2\pi}\left|\int x(u)h^*(u-t)e^{-j\,\omega\,u}du\right|^2. \qquad (6)$$

For discrete scales s, the wavelet filter $\psi(t)$ is applied along each temporal row of the spectrogram output:

$$P_x^{SP}(s,\zeta,\omega) \;=\; \frac{1}{s}\int P_x^{SP}(t,\omega)\psi^*\left(\frac{t-\zeta}{s}\right)dt\,. \qquad (7)$$

The energy across the wavelet translation axis ζ is integrated:

$$P_x^{SP}(s,\omega) \;=\; \int\left|P_x^{SP}(s,\zeta,\omega)\right|^2 d\zeta\,. \qquad (8)$$

The above equation yields a joint frequency representation with non-uniform resolution in the modulation frequency dimension, as indexed by the discrete scale s. Our past work [9] showed the advantage of using wavelet over Fourier bases for discriminating two distinct modulation frequencies when the dimension of their representations is the same.

2.4 Time and frequency translation

A practical classification system needs to be robust to changes in the signal. In this paper, we address the robustness of our analysis to time and frequency translation. If the signal is shifted in time, t_0, and in acoustic frequency, ω_0, i.e. $y(t) = e^{j\omega_0 t}x(t-t_0)$, then the joint frequency representation of the shifted signal can be expressed as

$$P_y^{SP}(\eta,\omega) \quad = \quad e^{-j\eta t_0}P_x^{SP}(\eta,\omega-\omega_0)\,. \qquad (9)$$

The effect of a time shift results in a phase shift in the η dimension of $P_x^{SP}(\eta,\omega)$. By using the absolute value of $P_y^{SP}(\eta,\omega)$, the estimate is invariant to the time shift. For the modulation scale representation $P_y^{SP}(s,\omega)$, if the time shift is relatively small compared to the integration time in Equation (8), then the estimate is approximately invariant to the effect:

$$P_y^{SP}(s,\omega) \quad \simeq \quad P_x^{SP}(s,\omega-\omega_0)\,. \qquad (10)$$

An acoustic frequency shift causes a shift in the ω dimension for both $P_y^{SP}(\eta,\omega)$ and $P_y^{SP}(s,\omega)$. We will discuss the approach for estimating features which are insensitive to this frequency shift in Section 3.3.

3. EXPERIMENTS

3.1 Signal Interception

There are many modulation types used to transmit digital messages over analog channels. In many applications such as interception of battlefield communications, modulation type is unknown, and identification of the type is a critical first step in monitoring of the communication channel. Current systems rely on the accuracy of a human listener. An operator manually demodulates an intercepted signal to the audible frequency range at which point an expert listener then identifies the modulation type. Past research into automatic identification of modulation type has used a combination of short-term spectral features, i.e. [2]. A weakness of features of this type is that they are sensitive to frequency translations which may be induced during the initial demodulation of the signal into the audible range by the human operator. Although it is possible to align the center frequencies of signals to eliminate this drawback, doing so requires a rather exact estimate of the carrier frequency [13]. We will show that our features derived from joint frequency analysis not only provide high discrimination but also low sensitivity to frequency shift.

3.2 Data Collection

The real communication signals used in the experiments were collected from http://rover.vistecprivat.de/~signals. The audio data was labeled by an expert listener afterward. There were a total 216 files including four different modulation classes, FSK, MFSK, PSK (phase shift keying and multilevel PSK), and MCVFT (multichannel FSK and PSK). Each file had different length and sampling rate. The number of files for each signal class is shown in Table 1.

Table 1: The number of communication signal files used in the experiments.

	FSK	MFSK	PSK	MCVFT
Number of files	77	41	72	26

3.3 Feature Extraction

Each file was resampled to 11025 Hz. After removing silences, the resampled audio was then windowed into 3-second blocks which overlapped by 2.75 seconds. A spectrogram was computed for each block using a Hanning window of length 128 samples and a window shift of 21 samples thereby reducing the subband sampling rate to about 512 Hz before the modulation transform. Biorthogonal wavelet filters, with 8 different dyadic scales, were used to produce one modulation frequency vector for each acoustic subband. The output features included modulation scale $P_{mod}[s,k]$, and spectral estimate $P_{spec}[k]$ features.

As previously discussed, the nature of the initial demodulation process may induce an acoustic frequency shift in P_{mod} and P_{spec}. Due to this frequency translation, we cannot directly apply P_{mod} and P_{spec} to typical classifiers. To remove the frequency sensitivity in P_{mod}, a singular value decomposition (SVD) was considered. Using the SVD, we can estimate the acoustic frequency vector, P^a, and modulation scale vector, P^m, given P_{mod}, the feature matrix with rank r, by Equation (11) (where σ is a nonnegative weight).

$$P_{mod}[s,k] \quad = \quad \sum_{j=1}^{r}\sigma_j P_j^a[k]P_j^m[s] \qquad (11)$$

Using Equation (10), the modulation scale features of the shifted signal can be approximated as $P_{mod}[s,k-k_0]$ where k_0 is amount of frequency translation. Due to the sparseness of the joint frequency representation, $P_{mod}[s,k-k_0]$ could be viewed as a vertical shift of the non-zero support region of Figure 1b and Figure 2b while the structure in the modulation frequency dimension, or horizontal axis in the plots, remains the same. This is equivalent to a row permutation of the P_{mod} matrix. It can be shown that a row-permutation in this matrix results in a row permutation of the left matrix in the SVD which implies that the frequency shift affects only P^a. Since P^m and σ were insensitive to acoustic frequency shift, they were chosen for our features. Because P_{mod} can be mostly represented using only one basis vector, we derived our new 8 dimensional modulation features as

$$\textit{modulation features} \quad = \quad sign\left(\sum_s P_1^m[s]\right)\sqrt{\sigma_1}P_1^m[s] \qquad (12)$$

For the P_{spec} vector, four scalar features which were also insensitive to frequency shift were extracted including entropy and bandwidth of P_{spec}, and the mean and standard deviation of the demodulated spectrum estimate. The normalized central second moment of the analytic signal [2] was also used for our features. We refer to these five features collectively as "spectral estimate" features. To reduce the dynamic range of the estimation, all features were normalized by the standard deviation estimated from all signal classes before classification.

3.4 Results

A probabilistic classifier, Gaussian mixture model (GMM), was used for our experiments. Due to the limited number of examples in each signal class, the method of withholding each example in turn for testing while training on all the rest ("leave-one-out" approach), was employed to evaluate the performance for the classifier. For a given multi-frame test file, each frame was assigned to one of the four modulation classes having the highest likelihood score. The modulation class winning the largest number of frame assignments was chosen as the class for that exam-

ple. Table 2 shows the accuracy results with different feature types and number of mixtures. The results show the improved performance when the number of mixtures is increased. Using short-term spectral estimate features exclusively provides performance comparable to using only modulation features. However, adding modulation features to spectral estimate features yielded markedly improved performance for all numbers of mixtures. The improved performance of adding modulation features also was seen for a *k*-nearest neighbor classifier.

Table 2: Percentage accuracy results using Gaussian mixture models

Features	Number of mixtures				
	2	***4***	***6***	***8***	***10***
modulation	51.9%	58.3%	62.0%	61.6%	65.7%
spectral	61.6%	60.7%	62.5%	66.2%	66.2%
combined	66.2%	69.9%	71.8%	78.2%	75.5%

4. CONCLUSIONS

This paper discussed joint frequency analysis for non-stationary signal classification. Our joint frequency representation has provided not only short-term information but also long-term information about the signal. Since it was designed using time-frequency theory and psychoacoustic results, the resulting approach has potential for a wide range of non-stationary signal types. Communication signal type classification was used for experimental verification. The experimental results using the probabilistic classifier showed that adding modulation features to spectral estimate features significantly improved the overall system performance. For the best performance, the relative improvement was 18% for a GMM classifier with 8 mixtures.

Acknowledgement

This work was supported by the Office of Naval Research and the Air Force Research Laboratory.

5. REFERENCES

[1] D. H. Kil and F. B. Shin, *Pattern Recognition and Prediction with Applications to Signal Characterization*. Woodbury, New York: AIP Press, 1996.

[2] J. S. Sewall and B. F. Cockburn, "Voiceband signal classification using statistically optimal combinations of low-complexity discriminant variables," *IEEE Transactions on Communications*, vol. 47, pp. 1623-7, 1999.

[3] R. K. Fish, *Dynamic models of machining vibrations, designed for classification of tool wear,* Ph.D. Dissertation, University of Washington, Seattle, 2001

[4] H. Hermansky, "Should recognizers have ears?," *Speech Communication*, vol. 25, pp. 3-27, 1998.

[5] ISO/IEC JTC1/SC29/WG11 MPEG, IS13818-7 "Information Technology - Generic Coding of Moving Pictures and Associated Audio, Part7: Advance Audio Coding," 1997.

[6] R. E. Crochiere, S. A. Webber, and J. L. Flanagan, "Digital coding of speech in sub-bands," Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, 233-6, 1976.

[7] M. S. Vinton and L. E. Atlas, "Scalable and progressive audio codec," Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing Proceedings, 3277-80, 2001.

[8] D. W. Massaro, "Preperceptual images, processing time, and perceptual units in auditory perception," *Psychological Review*, vol. 79, pp. 124-145, 1972.

[9] S. Sukittanon and L. E. Atlas, "Modulation frequency features for audio fingerprinting," Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, 1173-76, 2002.

[10] W. A. Gardner, "Exploitation of spectral redundancy in cyclostationary signals," *IEEE Signal Processing Magazine*, vol. 8, pp. 14-36, 1991.

[11] L. Cohen, *Time-Frequency Analysis*. Englewood Cliffs, NJ: Prentice Hall, 1995.

[12] S. Ewert and T. Dau, "Characterizing frequency selectivity for envelope fluctuations," *Journal of the Acoustical Society of America*, vol. 108, pp. 1181-96, 2000.

[13] A. Polydoros and K. Kim, "On the detection and classification of quadrature digital modulations in broad-band noise," *IEEE Transactions on Communications*, vol. 38, pp. 1199-211, 1990.

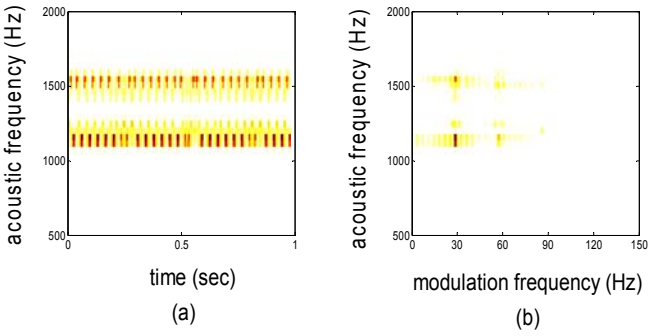

Figure 1: A spectrogram representation of FSK signal shows the change of two frequencies in (a) where the joint frequency representation of this signal exhibits a more compact representation in (b). More details are in Section 2.2.

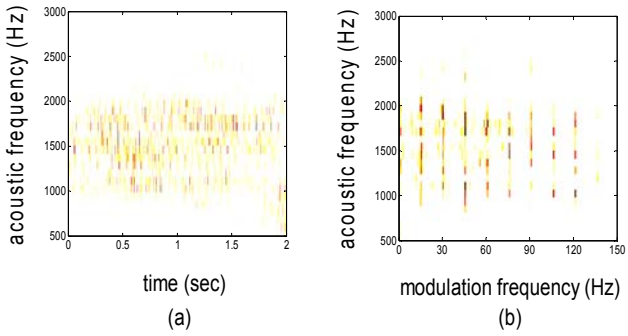

Figure 2: A spectrogram representation of PSK signal shows varying content in a wide range of frequency and time in (a) where the joint frequency representation of this signal exhibits a more compact representation in (b). More details are in Section 2.2.