Phoneme classification using the Hartley Phase Spectrum

Ioannis Paraskevas and Maria Rangoussi

Department of Electronics, Technological Educational Institute (T.E.I.) of Piraeus, Greece

<https://doi.org/10.36505/ExLing-2008/02/0046/000105>

Abstract

The phase of a signal conveys critical information for signal interpretation. However, signal phase extraction is not a straightforward process, mainly due to the discontinuities appearing in the phase spectrum. Previously, it was shown that for the application of audio classification, in case features of the processed phase spectrum are combined with magnitude related features they provide higher classification score compared to the case where features are extracted only from the signals' magnitude content. Moreover, it was shown that the Hartley Phase Spectrum encapsulates the phase content of the signals more efficiently compared to its Fourier counterpart. In this work, the importance of phase and the properties of the Hartley Phase Spectrum are demonstrated for the application of phoneme classification.

Key words: phoneme classification, Fourier Phase Spectrum, Hartley Phase Spectrum

Introduction

In this work we introduce the use of the Hartley Phase Spectrum (HPS) as an alternative to the standard Fourier Phase Spectrum (FPS) for the classification of speech signals. Accurate phase information extraction is critical for the success of the subsequent speech processing steps; yet difficulties in processing the phase spectrum of speech signals (Alsteris 2006) have led researchers to focus their investigation to magnitude based features only.

In speech signals (and other audio signals), the shape of the FPS is characterized by rapid changes, appearing as discontinuities, which are caused i) due to the use of the inverse tangent function, which is a highly discontinuous function ('extrinsic' discontinuities) (Tribolet 1977) and ii) due to properties of the signal itself ('intrinsic' discontinuities) (Al-Nashi 1989). The difficulty in processing the phase spectrum of signals, due to the aforementioned discontinuities, led the researchers to extract spectral features mainly from the magnitude spectrum and hence, ignore the phase related content of the signals (Paraskevas 2004).

In the FPS, the heuristic approach for the compensation of the 'extrinsic' discontinuities using the 'unwrapping' algorithm introduces ambiguities

which are more severe in the presence of even the lower possible noise level. The advantage of the HPS, compared to its Fourier counterpart, is that it does not convey ‘extrinsic’ discontinuities whereas its ‘intrinsic’ discontinuities can be easily compensated or removed when needed. The HPS, due to its structure, is less affected by the presence of noise; the noise immunity of the HPS, compared to its Fourier counterpart, is justified via the shapes of the respective probability density functions (Paraskevas 2008). The superiority of the HPS and its corresponding cepstrum, compared to its Fourier counterparts, has already been shown in signal analysis and pattern recognition (Paraskevas 2006 and 2007).

In this work, features are extracted from the processed phase spectrogram and are combined with features extracted from the magnitude spectrogram of the signals, for the application of phoneme classification. Specifically, we aim to classify four consonants namely: /h/ e.g. in “hill”, /s/ e.g. in “sister”, /f/ e.g. in “fat” and /th/ e.g. in “thin” (TIMIT® Acoustic-Phonetic Continuous Speech Corpus), based on their spectral characteristics. Similarly to our previous work - which involved audio signals (Paraskevas 2006) - the experimental results show that the classification score obtained in the case where the features extracted from the magnitude spectrograms are combined with the features extracted from the processed phase spectrograms of the phonemes is higher, compared to the case where the features employed are extracted only from their magnitude spectrograms.

Experimental set-up

The feature extraction step of the pattern recognition process includes calculation of statistical features from each spectrogram. Thus, each recording is represented by an $[8 \times 1]$ - sized feature vector. The statistical features employed here are chosen empirically as the most representative descriptive statistics: the variance, the skewness, the kurtosis, the entropy, the inter-quartile range, the range, the median and the mean absolute deviation.

In steps, each recording is divided into equal-length frames (256 samples) with zero-padding of the last frame if necessary, windowed with a Hanning window and transformed to the frequency domain. The transformed frames are placed row-wise in a matrix and hence the spectrograms are formed (Fourier and Hartley Phase Spectrograms, Fourier and Hartley Magnitude Spectrograms and Hartley Transform (Bracewell 1986) Spectrogram). Each of the spectrograms has to be presented to the classifier with its dimensionality reduced. Hence, the statistical features are calculated from each spectrogram for each recording, in order to compress the information into a compact feature vector. Then, from the aforementioned spectrograms, three were selected based on their recognition performance, namely: i) the

Fourier Magnitude (FM) Spectrogram, ii) the Hartley Phase Spectrogram (HPS) via the Discrete-Time Hartley Transform (DTHT) and iii) the Hartley Magnitude (HM) Spectrogram - all three spectrograms convey distinct information.

The classification decision for each recording is based on the 'majority vote' classification concept. For each recording, the statistical features are calculated for each of the three aforementioned spectrograms. Consequently, three feature vectors are formed for each recording. A phoneme is classified to a certain class based on the minimum distance between the test and the target vectors where the classifier used for each of the three independent 'experts' is the Mahalanobis (Mahalanobis 1936) and (Webb 2002). In case two or three out of the three independent 'experts' agree, then a recording is classified to this class. In case of tie, the decision is taken based on the Fourier Magnitude independent 'expert'. Each class of phonemes - /h/, /s/, /f/ and /th/ - consists of ten recordings where seven of them are used as training data and the rest as test data.

Experimental results

A series of two-class experiments between the aforementioned speech signals are carried out - Table 1 presents the correct classification scores, averaged across all six two-class experiments.

Table 1. Correct classification scores (%).

Feature	FM	FM, HPS via the DTHT & HM
Average scores (4 classes)	52.8	75.0

The second column of the table presents the correct classification score in the case where the features are extracted only from the Fourier Magnitude spectrogram and the third column presents the correct classification score of the combined scheme (Fourier Magnitude, HPS via the DTHT and Hartley Magnitude spectrograms). Similarly to the experiments involved the audio signals (Paraskevas 2006), the classification score is increased (22.2%) in the case where the phase-related features are also employed, compared to the case where features are extracted only from the Fourier Magnitude content of the speech signals. Moreover, as in (Paraskevas 2006), the experimental results show that the classification rate is higher in the case where i) the statistical features are extracted from the Hartley rather than the Fourier Phase spectrogram, ii) the discontinuities are compensated, iii) the 'difference' is used when estimating the phase spectrum.

Conclusions

This work - focused on the feature extraction stage of the pattern recognition process - shows the importance of phase as a feature extraction tool for phoneme classification. The experiments involving speech signals show that the proposed phase-related features, when combined with conventional features based on the magnitude spectrum, form a significantly more discriminative feature set. Moreover, the experimental results indicate that the features extracted from the Hartley Phase Spectrum, present the information content of the signals to the classifier in an improved manner, as compared to the Fourier Phase case. As a conclusion, features extracted from the Hartley Phase Spectrum, if employed along with conventional magnitude related features, provide a significantly improved feature set for frequency-domain statistical classification.

References

- Al-Nashi, H. 1989. Phase Unwrapping of Digital Signals. *IEEE Transactions on Acoustics, Speech and Audio Processing* 37 (11), 1693–1702.
- Alsteris, L.D., and Paliwal, K.K. 2006. Further intelligibility results from human listening tests using the short-time phase spectrum. *Speech Communication*, 48, 727-736.
- Bracewell, R. N. 1986. The Fourier Transform and Its Applications. In 2nd ed. chap. 19, McGraw-Hill, New York.
- Mahalanobis, P.C. 1936. On the generalized distance in statistics. *Proceedings of the National Institute of Science of India*, vol. 12, pp. 49-55.
- Paraskevas, I., and Chilton, E. 2004. Combination of Magnitude and Phase Statistical Features for Audio Classification. *Acoustics Research Letters Online*, vol. 5, Issue 3.
- Paraskevas, I., Chilton, E., and Rangoussi, M. 2006. Audio Classification Using Features Derived From The Hartley Transform. *Proc. of the 13th Int. Conference on Systems, Signals and Image Processing*, Budapest, Hungary, pp. 309-312.
- Paraskevas, I., and Rangoussi, M. 2007. The Hartley Phase Cepstrum as a Tool for Signal Analysis. *Advances in Nonlinear Speech Processing*, Springer Lecture Notes in Computer Science, vol. 4885, pp. 204 - 212.
- Paraskevas, I., and Rangoussi, M. 2008. The Hartley Phase Spectrum as a noise-robust feature in speech analysis. To appear in *Proc. of the ISCA Tutorial and Research Workshop (ITRW) on Speech Analysis and Processing for Knowledge Discovery*. Aalborg, Denmark.
- Tribolet, J. 1977. A new phase unwrapping algorithm. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 25, no. 2, pp. 170 – 177.
- Webb, A. 2002. *Statistical Pattern Recognition*. In 2nd ed., John Wiley and Sons, England.