Foreground Harmonic Noise Reduction For Robust Audio Fingerprinting

Matthew C. McCallum, Gracenote {matthew.mccallum@nielsen.com}

Overview

- Many audio fingerprinting algorithms rely on spectral peaks as unique features with which to identify an audio signal.
- Spurious spectral peaks often arise in acoustic audio fingerprinting applications due to speech, humming and/or singing.
- Independently, spurious peaks are largely indistinguishable from desirable peaks, however, groups of these peaks often have distinctly different features from the audio of interest, e.g., music.
- By searching for outliers in pitch contour characteristics, spurious peaks may be identified and removed prior to any fingerprint processing.
- This technique is more efficient than typical audio source separation algorithms that might address the same issue (such as non-negative matrix factorization or deep learning based approaches).

Contour Tracing

Contour tracing is performed analogous to previous literature with the additional consideration of phase.

Given a Fourier Transform:

$$X[k,m] = \sum_{n=mM}^{mM+N-1} x[n]\, w[n-mM]\, e^{\frac{-i2\pi nk}{K}}$$

The following quantities are calculated for each bin:

Magnitude: $A_{k,m} = \frac{2|X[k,m]|}{|W(\omega_{k,m})|}$

Frequency: $\omega_{k,m} = \frac{2\pi k}{K} + \frac{\left(\angle X[k,m] - \angle X[k,m-1] - \frac{2\pi MK}{K}\right) \bmod 2\pi}{M}$

Phase: $\phi_{k,m} = \angle X[k,m] + \angle W(\omega_{k,m})$

Peaks with low magnitude or inconsistent frequency are filtered out whilst the remainder are grouped into contours and harmonic sets ensuring the following between time frames:

Magnitude Continuity: $\frac{A_{k,m}}{A_{root,m}} > \Delta_A$

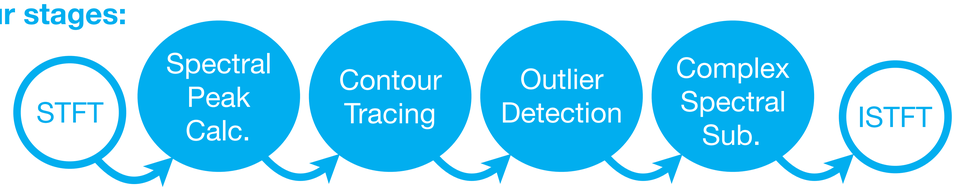
Frequency Continuity: $|\omega_{s,m-1} - \omega_{k,m}| < \Delta_f$

Phase Continuity: $\min_{\psi} |(\psi 2\pi - \phi_{k,m} + \phi_{s,m-1} + \omega_{s,m-1} M) \bmod 2\pi| < \Delta_\phi$

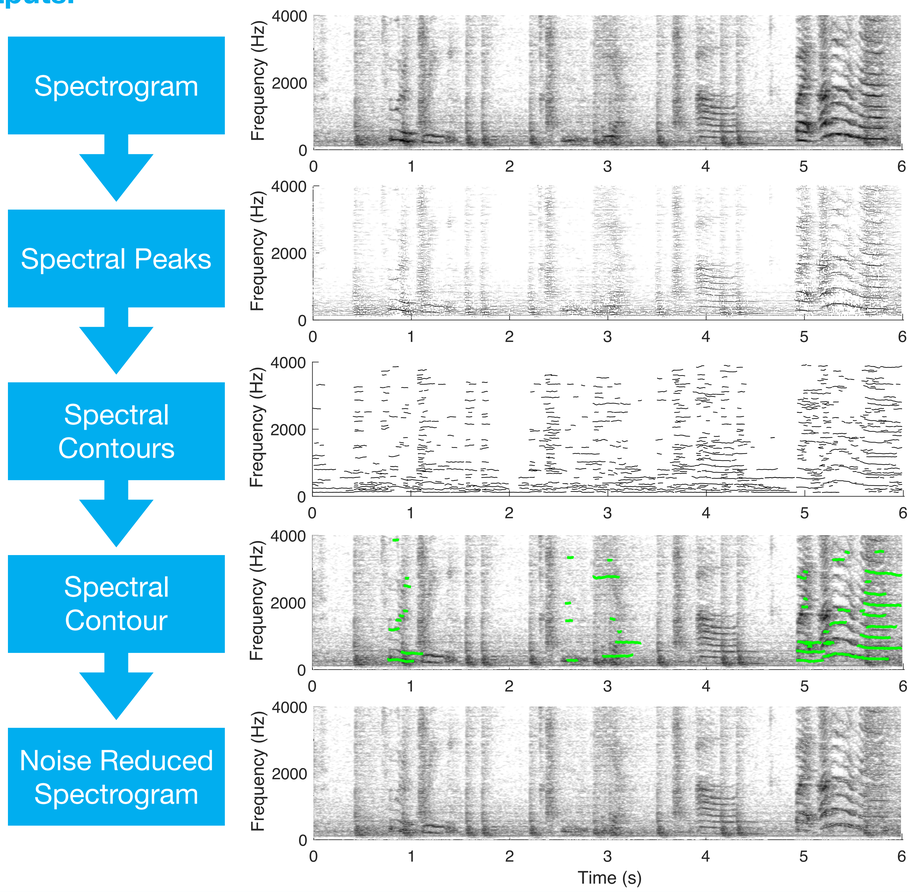
gracenote. A NIELSEN COMPANY

System Overview

Four stages:



Outputs:



Contour Characteristics

While individual spectral peaks in music and speech are hard to distinguish from one another. Groupings of these peaks across time and frequency have distinct properties.

First, the following are used to filter out spurious contours:

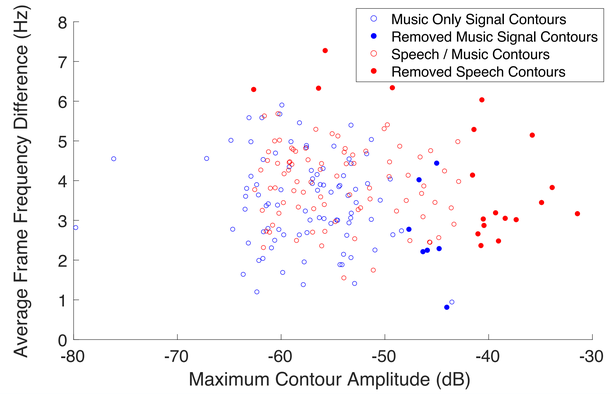
- Contour length
- Contour tracing signal to noise ratio

Next, the following are used to identify unusual / non-musical contours:

- Average absolute change in frequency
- Maximum contour amplitude

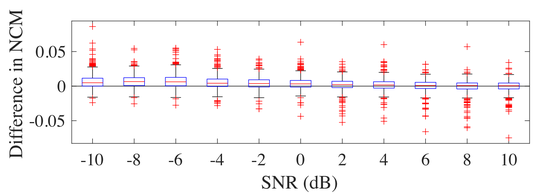
Outliers are selected based on a distance from the sample median.

Contours that are greater than a multiple of the median absolute deviation are removed from the signal via complex spectral subtraction.

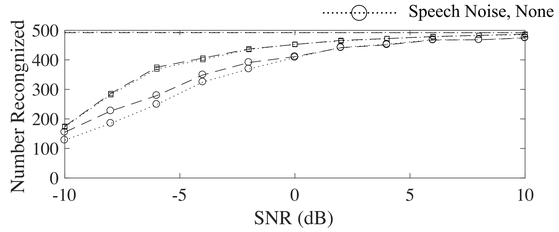


Results

While the objective is not to improve listening quality of the audio, it can be seen that the spectrum better represents the speech free audio spectrum by the Normalized Covariance Measure (NCM):



Harmonic Noise Reduction (HNR) improves performance in the presence of speech noise but maintains performance with wideband noise sources like babble.



A significant number of previously unrecognized fingerprints are now recognizable using Gracenote's latest fingerprinting algorithm thanks to Harmonic Noise Reduction (HNR).