The Envelope Modulation Spectrum and Its Application in Rhythm Studies

Susan LeGendre*, Julie Liss** & Andrew Lotto*

*Department of Speech, Language and Hearing Sciences, University of Arizona
**Motor Speech Disorders Laboratory, Arizona State University

Envelope Modulation Spectrum (EMS) Concept

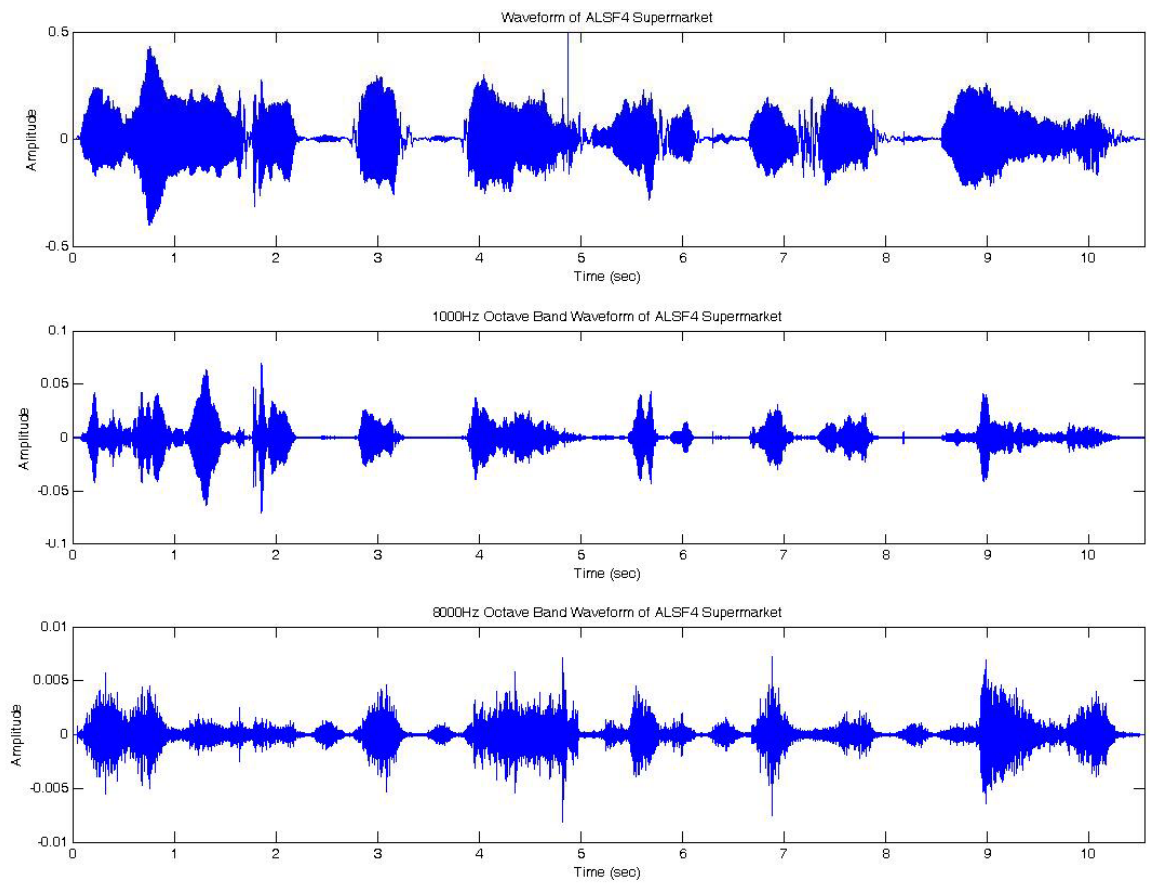
- The envelope modulation spectrum is a representation of the slow amplitude modulations in a signal
  - Steeneken and Houtgast (1980) used the envelope spectrum of a full signal to quantify effects of room acoustics on speech. It is an acoustic measure that makes no assumptions about the linguistic structure in the signal
- It is equivalent to the power spectrum of the signal's amplitude envelope
  - Depicts the distribution of energy in the amplitude fluctuations across frequencies
- Has since been proposed to be useful in analyzing the rhythmic structure of speech [Drullman, Festen, & Plomp (1994); Greenberg, Arai, & Silipo (1998)]
  - Intelligibility of speech may directly correlate with the integrity of the EMS amplitude in the 3-8Hz region
    - The importance of this region is thought to extend across languages
- We have taken the idea of the modulation spectrum and applied it to octave bands to provide a fuller analysis of modulations in the signal.

EMS Program

- In general, EMS program designed to:
  - Extract amplitude envelope from imported .wav file
  - Calculate the modulation/power spectrum for the .wav file
  - Output a matrix of dependent variable calculations and related figures
- It does this for the entire signal, and 7 octave bands:
  - 125Hz
  - 250Hz
  - 500Hz
  - 1000Hz
  - 2000Hz
  - 4000Hz
  - 8000Hz

EMS Program Filtering Example

- Waveforms of original signal and two octave bands

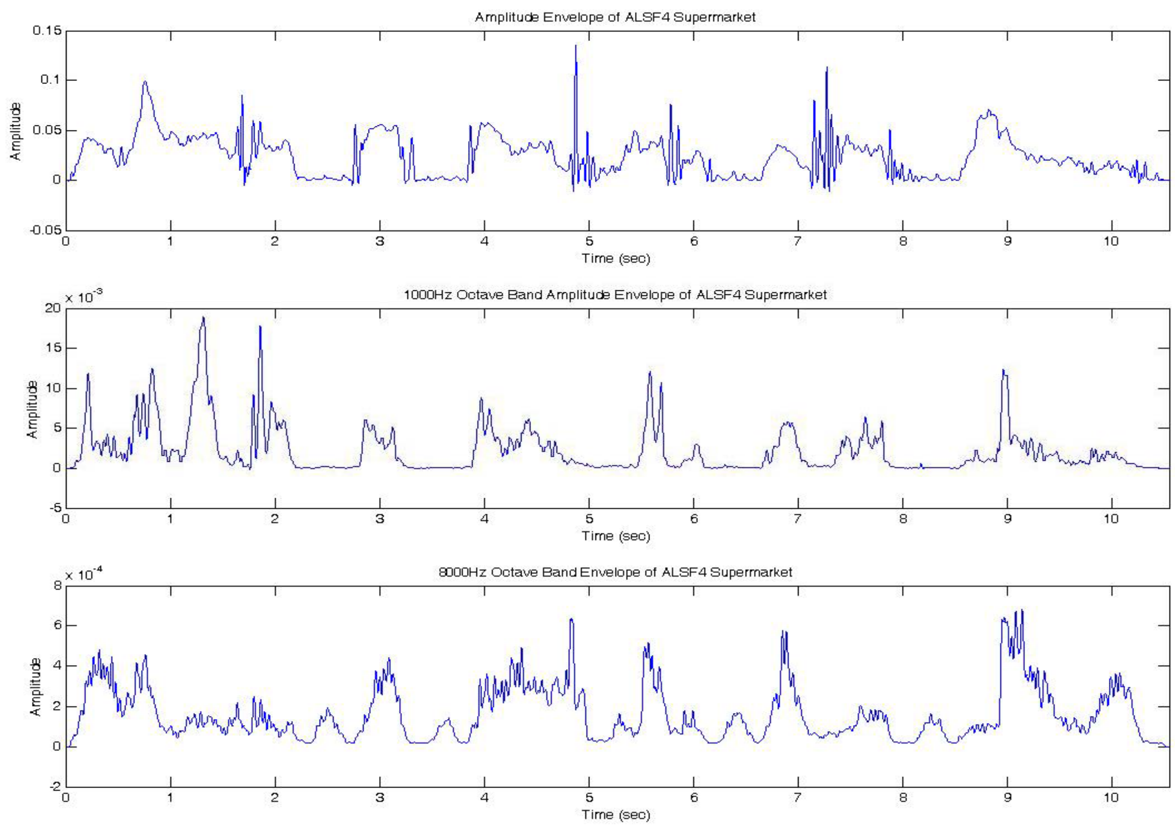


EMS Program Specifics: Envelope Extraction

- The amplitude envelope of the entire signal and each octave band was extracted using a 4^{th} order Butterworth low-pass filter with 30Hz cutoff frequency
  - Signal was half-wave rectified before applying filter

EMS Program Envelope Extraction Example

- Amplitude envelopes of original signal and two octave bands

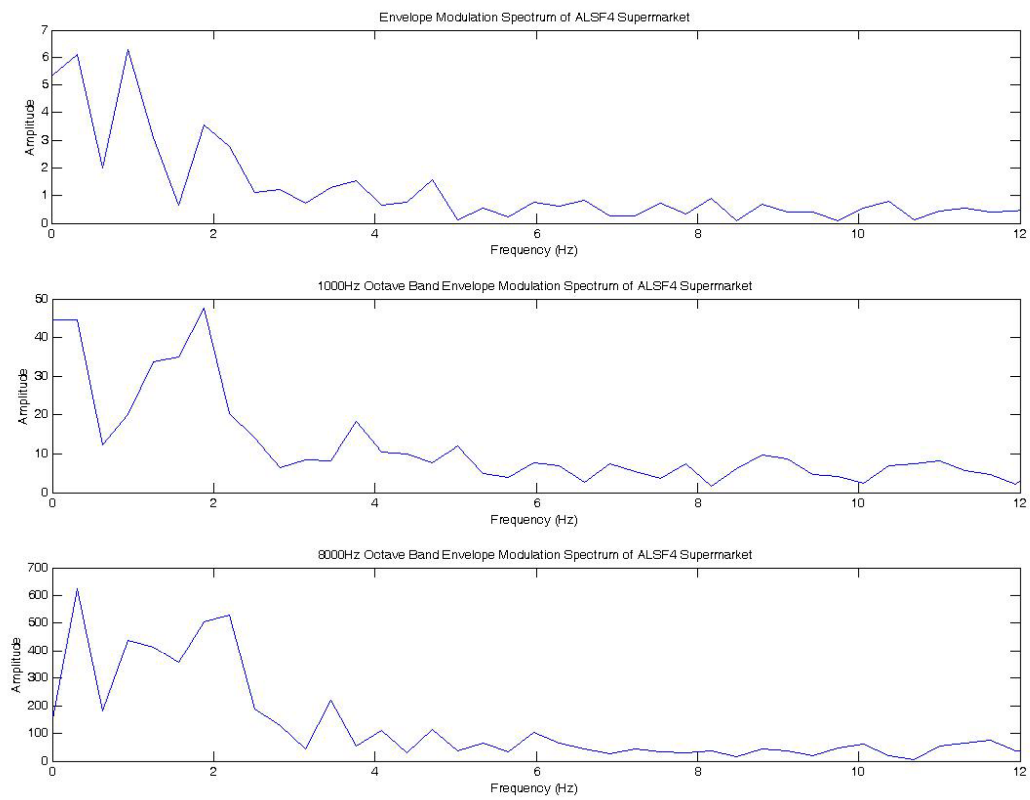


EMS Program Specifics: Modulation Spectra

- Modulation/Power spectrum calculated from energy in extracted amplitude envelopes for original signal and each octave band
- Before calculating power spectrum, the signals were
  - Downsampled to 80Hz
  - Windowed with a Tukey Window
  - Mean of signal subtracted
- Power spectrum calculated by taking the FFT of the signal's magnitude, squared
  - Resulting signal then normalized to maximum autocorrelation

EMS Program Modulation Spectra Example

- Modulation spectra of original signal and two octave bands



EMS Program Specifics: Dependent Variable Calculations

- Six dependent variables were derived from the full signal EMS and for each of the octave bands (48 total variables)
  - Peak_Freq = Frequency in Hz of highest peak
  - Peak_Amp = Amplitude of highest peak normalized by dividing average amplitude in spectrum
  - E3-6Hz = Average energy between 3-6 Hz normalized.
  - Below40 = Summed energy (normalized) below 4 Hz
  - Above40 = Summed energy (normalized) above 4 Hz
  - Ratio40 = Ratio of Below40/Above40

EMS Program Dependent Variable Calculations Example

- Dependent variable spreadsheet for sentence used in all previous examples

	Peak_Freq	Peak_Amp	E3_6Hz	Below40	Above40	Ratio40
Entire Signal	0.943315508	21.82486652	2.848423391	74.91199219	23.32032862	3.212304313
125Hz Octave Band	0	20.30115774	4.857927807	55.50631309	43.26055685	1.283069778
250Hz Octave Band	0.943315508	32.48294794	2.992968195	118.2830426	22.12538814	5.346032433
500Hz Octave Band	0.943315508	30.5146054	3.430702422	120.7635534	19.57164458	6.170332435
1000Hz Octave Band	1.886631016	20.3271818	3.927389572	82.38787972	34.24232569	2.406024651
2000Hz Octave Band	1.257754011	21.87931945	5.017092935	83.71526295	38.50769392	2.17398796
4000Hz Octave Band	0.943315508	19.64429971	4.004329313	70.77600894	35.86167066	1.973583707
8000Hz Octave Band	0.314438503	20.04194291	2.610194179	77.32286723	19.95912175	3.874061604

Dysarthrias in General

Darley, Aronson, & Brown (1969); Liss, et al. (2009)

- Group of related motor speech disorders resulting from neurologic lesions of the Central- or Peripheral Nervous System.
- Classifiable by acoustic-perceptual characteristics and underlying neuropathology.
- Severity and associated acoustic-perceptual characteristics highly variable from person to person
- Liss et al. (2009) were able to classify dysarthric speakers based on segment-based rhythm metrics. Is it possible to classify dysarthrias as accurately using EMS program?

Classifying Dysarthrias

- 5 sentences were spoken by talkers diagnosed with each type of dysarthria.
- EMS variables calculated from each sentence and then averaged within speaker
- Stepwise discriminant analysis conducted to classify each dysarthric type versus all other types.
- Only variables that significantly increase Wilks' lambda were included

Ataxic Dysarthria

Darley, et al. (1969); Liss, et al. (2009)

- Acoustic-Perceptual Characteristics
  - Imprecise Consonants
  - Excess and Equal Stress
  - Impaired Prosody
  - Irregular Articulatory Breakdown
- Most Predictive Variable(s) from Liss, et al. (2009)
  - VarcoC, rPVI-VC, nPVI-V ⟶ 85% discrimination
- Most Predictive Variable(s) from EMS Program
  - Below40_all, Peak_Freq_500Hz, Below40_1000Hz ⟶ 85% discrimination

Mixed Flaccid-Spastic Dysarthria

Darley, et al.(1969); Liss, et al. (2009)

- Example Neuropathology: Amyotrophic Lateral Sclerosis (ALS)
  - Progressive degeneration of both upper and lower motor neurons.
- Acoustic-Perceptual Characteristics
  - Imprecise Consonants
  - Hypernasality
  - Strangled Voice Quality
  - Slow Rate
  - Prolonged Syllables
- Most Predictive Variable(s) from Liss, et al. (2009)
  - VarvoV ⟶ 82% discrimination
- Most Predictive Variable(s) from EMS Program
  - Peak_Amp_8000Hz, Above40_1000Hz ⟶ 82% discrimination

Hyperkinetic Dysarthria

Darley, et al.(1969); Liss, et al. (2009)

- Example Neuropathology: Huntington's Chorea
- Acoustic-Perceptual Characteristics
  - Imprecise Consonants
  - Irregular Pitch and Loudness Changes
  - Strained-Strangled Voice Quality
  - Irregular Articulatory Breakdown
  - Irregular Rate Changes Across Syllable
- Most Predictive Variable(s) from Liss, et al. (2009)
  - VarcoC ⟶ 85% discrimination
- Most Predictive Variable(s) from EMS Program
  - Peak_Amp_250Hz, Above40_250Hz ⟶ 88% discrimination

Hypokinetic Dysarthria

Darley, et al.(1969); Liss, et al. (2009)

- Example Neuropathology: Parkinson's Disease
- Acoustic-Perceptual Characteristics
  - Monopitch
  - Increased Articulation Rate/ Rushes of Speech
  - Imprecise Articulation
  - Breathy Voice Quality
- Most Predictive Variable(s) from Liss, et al. (2009)
  - Articulation Rate ⟶ 100% discrimination
- Most Predictive Variable(s) from EMS Program
  - Below40_8000Hz, E3_6Hz_2000Hz, Ratio40_500Hz ⟶ 100% discrimination

Summary and Conclusions

- The results obtained from the EMS program indicate that it is clearly measuring something in the signal that is a reliable dysarthria predictor
  - In some cases, the most important variables for dysarthria discrimination occurred in the 8000Hz octave band (5680-11,360 Hz)
- Pros of using EMS instead of traditional rhythm metrics
  - No need to identify vowels and consonant intervals (or to make any linguistic presumptions)
  - Completely automated in MATLAB
  - Can take into account pauses and non-phonetic elements that may occur in the sample

References

Darley, F.L., Aronson, A.E., & Brown, J.R. (1969). Differential diagnostic patterns of dysarthria. *Journal of Speech, Language, and Hearing Research, 12, 246-269.*

Drullman, R., Festen, J.M. & Plomp, R. (1994). Effect of temporal envelope smearing on speech reception. *Journal of the Acoustical Society of America*, 95: 1053-1064.

Greenberg, S., Arai, T. & Silipo, R. (1998). Speech Intelligibility Derived from Exceedingly Sparse Spectral Information. *Proceedings of the International Conference on Spoken Language Processing, Sydney, December 1-4, 1998.*

Liss, J.M., White, L., Mattys, S., Spitzer, S., Lansford, K., Lotto, A.J., and Caviness, J. (2009). Classifying Dysarthrias by Speech Rhythm Metrics. *Auditory Cognitive Neuroscience Society(ACNS) Conference 2009.*

Steeneken, H. & Houtgast, T. (1980). A physical method for measuring speech-transmission quality. *Journal of the Acoustical Society of America*, 67(1): 318-326.