



CLASSIFICATION OF EMOTION USING SUBAUDIBLE FREQUENCIES IN VOCAL DATA

Cody Narber
A Master's Thesis
Kansas State University
May 2008

Outline:

- Background Information
- Previous Work
- Data Collection
- Data Extraction
- Data Analysis
- Results
- Conclusions

Background:

Fourier Theorem: Every signal can be expressed as a sum of sine and cosine terms.

Using these terms, we can solve a variety of problems:

- Voice Recognition
- Voice Synthesis
- Voice Stress Detection
- Voice Emotion Analysis

There are special attributes of a voice sample that can be found using these terms.

Background:

Fundamental Frequency:

The fundamental frequency (F_0) is the perceived pitch of a particular waveform, and is sometimes known as the first harmonic frequency. This is typically the frequency that has the strongest amplitude.

Formants:

Formants are frequencies that are amplified along with F_0 from the reflections of sound that occur inside the oral cavity, nasal cavity, and vocal tract. Since these are affected by sound chamber of the body, different vowels are signified by different formants.

Background:

Parts of the Signal:

There is 3 major parts to a voiced signal: silence, voiced, and unvoiced. These are used for detecting parts of speech: separation between words, consonants, and vowels respectively.

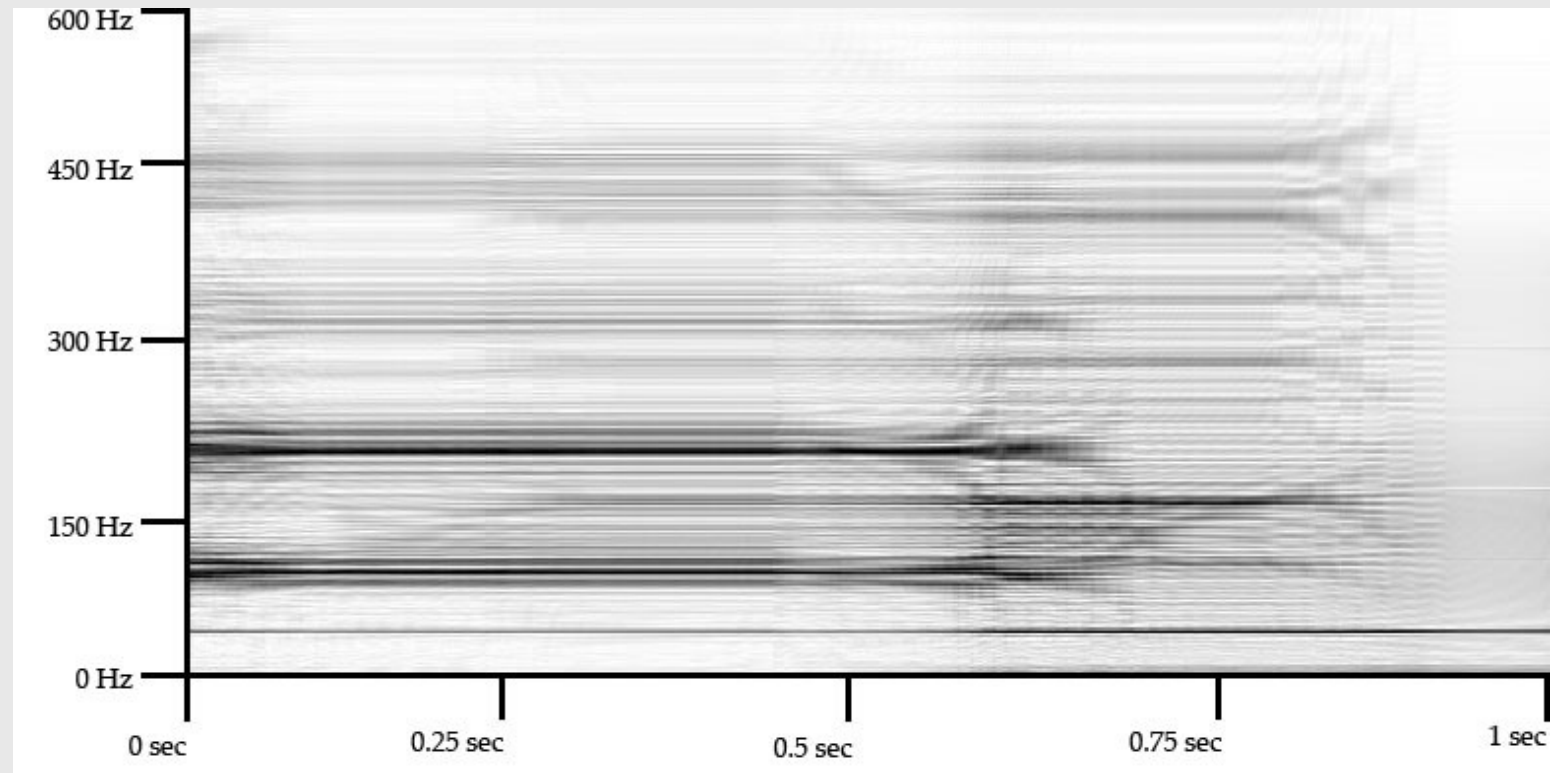
Temporal Properties:

Tempo is the speech rate, determined by words per time. Pauses are defined by abnormalities in the speech rate.

Background:

Sample Spectrogram:

A Spectrogram is a 3D plotting of *Amplitude of Frequencies over Time*.



"Dessert" Spectrogram

Previous Work:

Data Collection:

There exists 3 factors which has influenced sample gathering:

- Spontaneous vs. Elicited
- Classes of Emotions
- Acoustic Attributes Examined

Previous Work:

Elicited:

- Control over specific emotion
- Voice Stress Analysis
- Actors
- Movies, Sounds, Software

Spontaneous:

- Real Life Reactions
- Multimodal (Video/Audio)
- Expert Labeling
- TV chat shows, Interviews, Non-intrusive Surveillance

Previous Work:

Classes of Emotion:

- Negative vs. Positive
- Agitated vs. Passive
- Emotion Degrees (Hot/Cold)
- Emotion & Other States
- Subset of Emotions

Acoustic Attributes:

- F0 (mean, std, range)
- Energy
- Temporal
- Formants

Previous Work:

Results:

- Fundamental Frequency
- Energy
- Temporal changes
- Supported by both Elicited and Spontaneous

Data Collection:

Capture Statistics:

- Elicited with Software
- Neutral, Happy, Sad, Anger
- F0, Energy, Tremor Energy
- 8 Male Speakers w/ Limited Theater Experience
- NT3 Throat Microphone by IASUS Concepts Inc.
- Bose® Around Ear Headphones
- Stereo, 44100 Hz, 16 Bit res PCM Wave

Audio Capture Program

Data Extraction:

Information Extraction Steps:

- STFT Taken:
 - Window = 44100 Samples
 - Shift = 44 Samples
- Energy Smoothed with 100 Hz Window
- Local Maximum Found
- If highest peak was statistical insignificant (noise)
- Find Highest Peak in 50 – 250 Hz Range (F0)
- Record Amplitude of F0 (Amp)
- Sum all Energies in 8-12 Hz Range (Tremor Energy)
- Mean and Standard Deviation Calculated From All Three

Analysis Program

Data Analysis:

WEKA Clusterings:

- F0 (Average and Standard Deviation)
- Amp (Average and Standard Deviation)
- TE (Average and Standard Deviation)
- F0 and Amp
- F0, Amp, and TE Average
- F0, Amp, and TE Standard Deviation
- F0, Amp, and TE

Results:

Confusion Matrices of the Three Sets of Attributes:

F0 Clustering				
Neutral	Happy	Angry	Sad	
5	1	1	1	Neutral
2	4	0	2	Happy
4	1	2	1	Angry
4	1	1	2	Sad
Incorrectly Classified Clusters:				
			19	59.38%

Amp Clustering				
Neutral	Happy	Angry	Sad	
3	0	2	3	Neutral
0	4	2	2	Happy
1	2	1	4	Angry
1	0	1	6	Sad
Incorrectly Classified Clusters:				
			18	56.25%

TE Clustering				
Neutral	Happy	Angry	Sad	
3	0	1	4	Neutral
2	4	0	2	Happy
2	2	2	2	Angry
2	1	0	5	Sad
Incorrectly Classified Clusters:				
			18	56.25%

Results:

Confusion Matrices of the Addition of Tremor Energy Measures:

F0, Amp Clustering				
Neutral	Happy	Angry	Sad	
1	1	3	3	Neutral
3	5	0	0	Happy
1	1	4	2	Angry
0	1	0	5	Sad
Incorrectly Classified Clusters:				
			17	53.13%

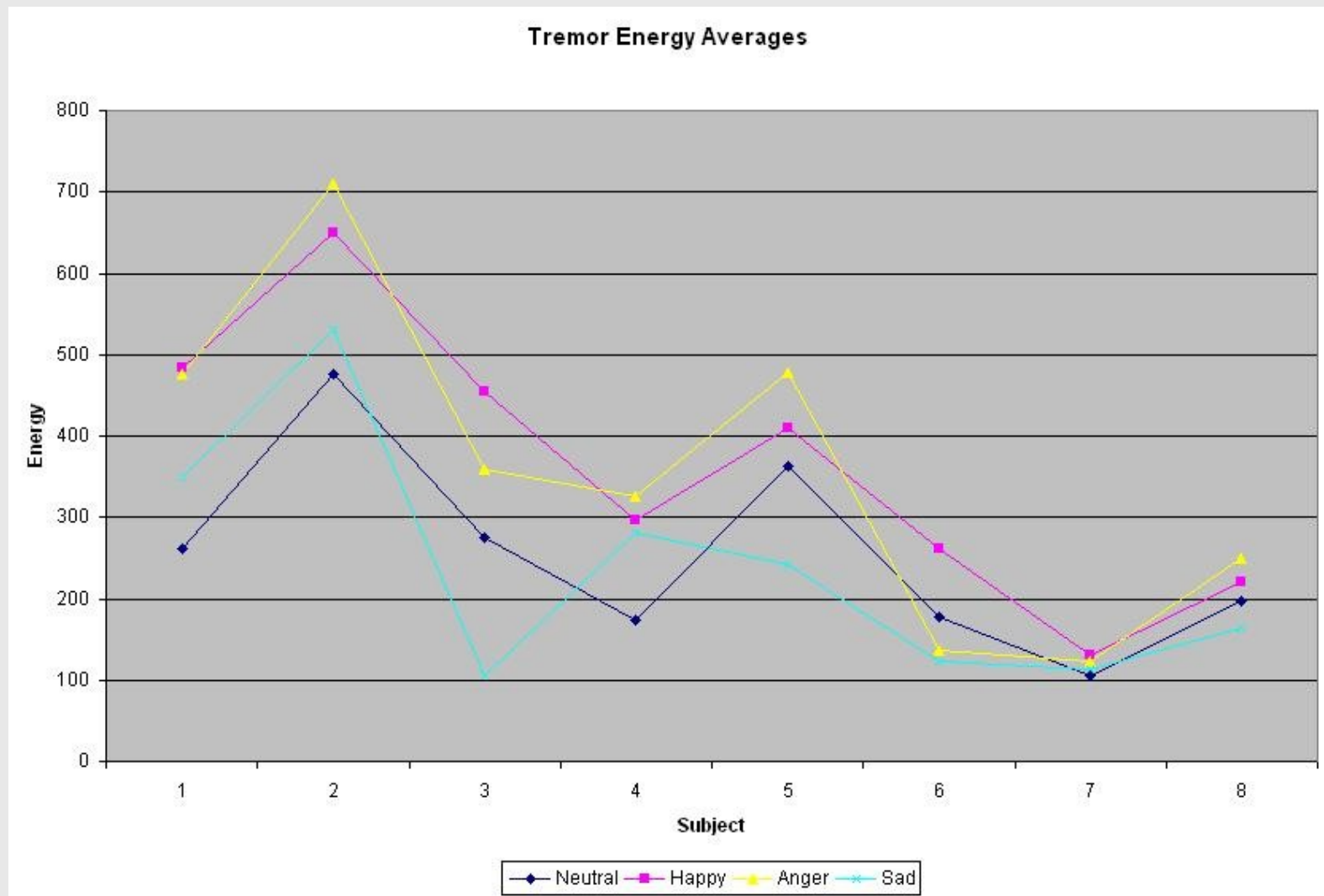
F0, Amp, aTE Clustering				
Neutral	Happy	Angry	Sad	
4	1	1	2	Neutral
0	6	2	0	Happy
2	1	3	2	Angry
3	0	0	3	Sad
Incorrectly Classified Clusters:				
			16	50.00%

F0, Amp, sTE Clustering:				
Neutral	Happy	Angry	Sad	
1	1	1	5	Neutral
4	4	0	0	Happy
2	1	4	1	Angry
1	0	1	6	Sad
Incorrectly Classified Clusters:				
			17	53.13%

F0, Amp, TE Clustering:				
Neutral	Happy	Angry	Sad	
5	1	1	1	Neutral
0	4	3	1	Happy
4	1	1	2	Angry
5	0	1	2	Sad
Incorrectly Classified Clusters:				
			20	62.50%

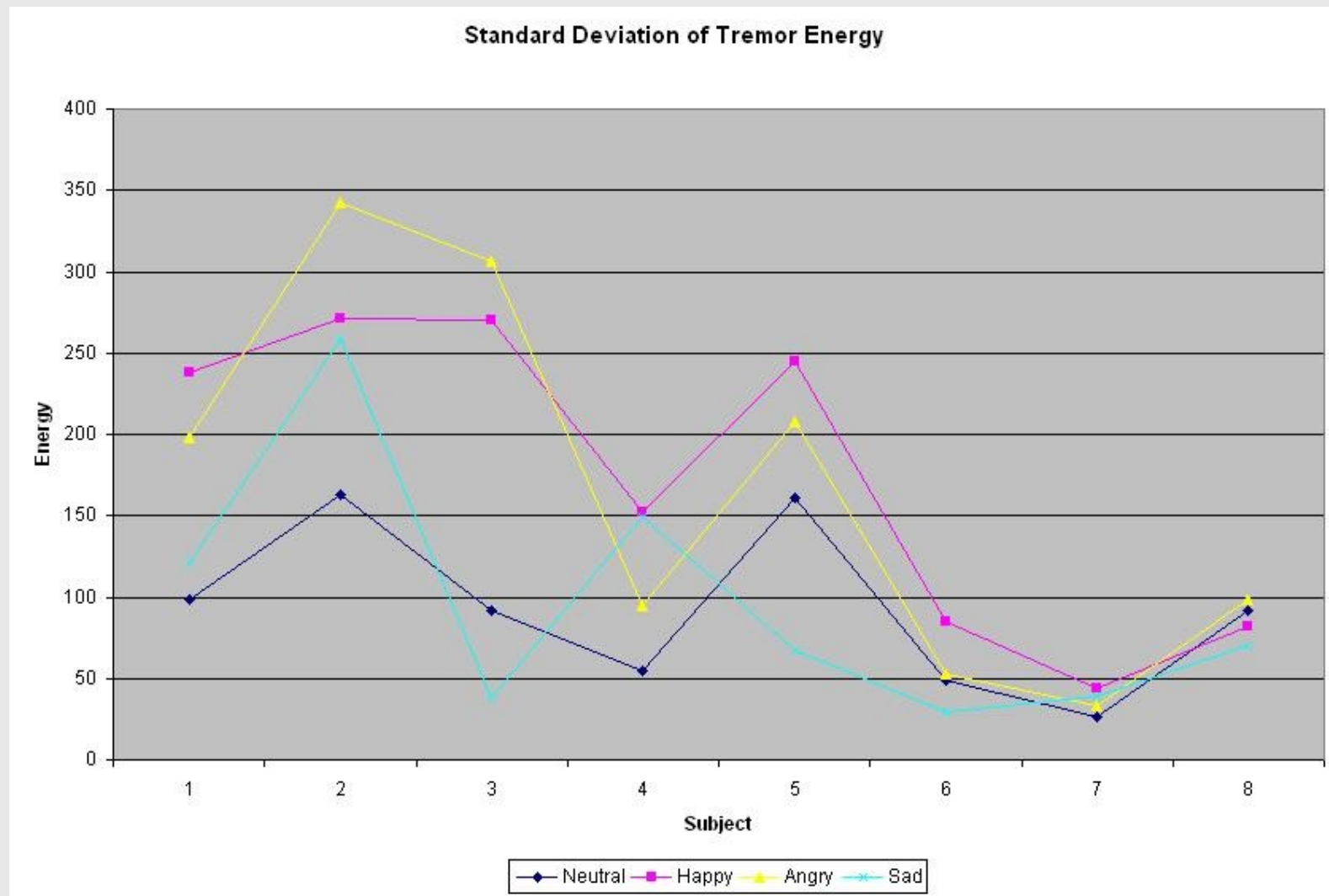
Results:

Plotting of Tremor Energy Averages for Each Speaker:



Results:

Plotting of Tremor Energy Variation in Each Speaker:



Conclusion:

Conclusions:

- Tremor Energy is a viable Attribute:
 - Shows the Agitation State Clearly
 - Improves Accuracy

Future Work:

- Larger Subject Space
- Actor & Non-actor
- Incentives for Subjects
- Larger Emotion Space
- Addition of Other Attributes



QUESTIONS?

QUESTIONS

???