



Automatic Mood Quantification of Contemporary Music

Chris Cooke

CS-714 – Capitol College

Professor Karim Chichakly

April 11, 2006



Music Information Retrieval

- Hard to find a suitable song in large database
- Existing methods tend to use
 - Metadata
 - Prior knowledge of database holdings
 - Marketing information

Wouldn't it be nice...

- **Could we retrieve songs based on mood?**
 - Requires no prior knowledge of holdings by requester
 - Doesn't tie play list to genre or artist
 - Matches one way people attempt to find songs
 - Possible additional benefit of targeted mood induction

Contents

- Current state of the art
- A different approach
- Overview of the system
- Evaluation
- Discussion
- Future research

Current state of the art

■ Current mood-based approaches

- Focus on classification rather than quantification
- Some rely on human classification rather than automatic classification

■ Classification

- Prevents mixture and degree of mood
- Crisp results not as useful in fuzzy expert systems

■ Human classification

- Not all songs evaluated
- More popular songs evaluated more
- New songs have no rating until a human evaluates

A different approach

- Use a machine learning algorithm to quantify mood parameters according to a model
- Thayer's model
 - Two dimensions: energy, tension
 - Can represent as a planar segment
- By quantifying on a plane segment
 - Could provide input to a fuzzy expert system
 - Could trace a path from one mood to another to generate a mood induction play list

Overview of system

- Thayer's model
- Sound clips
- Human evaluation
- Machine evaluation
 - Feature extraction
 - Machine learning algorithm

Thayer's model

- Two dimensions

- Energy – getting up and doing something versus sitting at rest
- Tension – mental turmoil versus peaceful calm

- Energy and tension are **NOT** independent

Sample music clips

- Call for clips posted to a forum
- 20 seconds
- Uniform mood throughout clip
- No limit on number of clips per artist
- MP3 format specified, but not parameters

Human evaluation

- Activation/Deactivation Adjective Check List (AD ACL)
- Online survey
 - 50 clips
 - Short (20 adjective) evaluation per clip
- Rescaled for 0-100 result

Sample Survey Page

- 50 clips
- 20 adjectives
- Same order of clips and adjectives for each participant
- No time limit
- Survey could be saved and resumed

Music Mood Recognition Survey

http://geekido.org/survey.php

Getting Started Latest Headlines MacIDOL Top 40 MacIDOL Apple Yahoo! Naked News News Music

MUSIC MOOD RECOGNITION SURVEY

PAGE 2 OF 51

QUESTIONS MARKED WITH A * ARE REQUIRED.

CLIP 1

*1. PLEASE CHOOSE THE DEGREE TO WHICH EACH ADJECTIVE DESCRIBES THE CLIP.

	Definitely describes	Slightly describes	Cannot decide	Definitely does NOT describe
active	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
placid	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
sleepy	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
jittery	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
energetic	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
intense	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
calm	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
tired	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
vigorous	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
at-rest	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
drowsy	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
fearful	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
lively	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
still	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
wide-awake	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
clutched-up	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
quiet	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
full-of-pep	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
tense	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Done

What musical features contribute to mood?

■ Pitches

- Lower pitches “darker”, more tension, less energy
- Ratios of pitches interpreted as dissonance → tension
- Chord progressions go from tension to resolution

■ Harmonics

- Pure tones calmer
- More harmonics → more tension

■ Beat

- Faster tempo → more energy, more tension
- Irregular beats → more tension

■ Volume

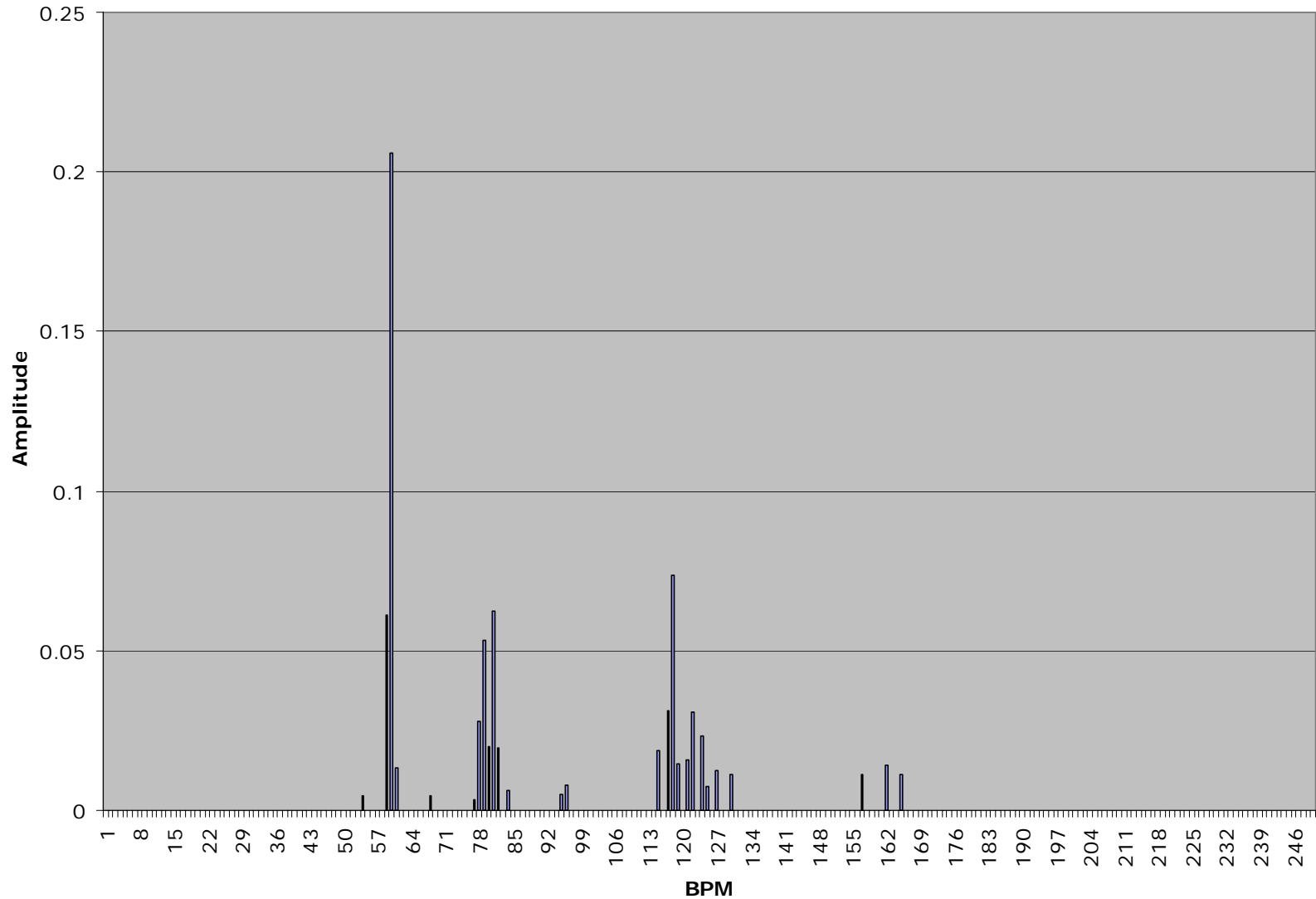
- More volume → more energy

Audio Features for Music

- Short-term Fourier Transform
- Mel-frequency Cepstral Coefficients
- Time Domain Zero Crossings
- Beat Histogram
- Pitch Histogram

Beat Histogram Example

Clip 4 Beat Histogram



Final feature set

■ STFT (6 features)

- Mean and standard deviation of spectral centroid, spectral rolloff, spectral flux

■ MFCC (10 features)

- Mean and standard deviation of first 5 MFCCs

■ Zero Crossings (2 features)

- Mean and standard deviation

■ BH (8 features)

- Amplitude and position of largest 2 peaks
- Ratio of peaks
- Sums from 40-90 BPM, 90-140 BPM, and 40-250BPM

Note: STFT, MFCC, and Zero Crossings used 23 ms analysis windows averaged over 1 second texture window. Final features are averages of texture window values over the entire clip.

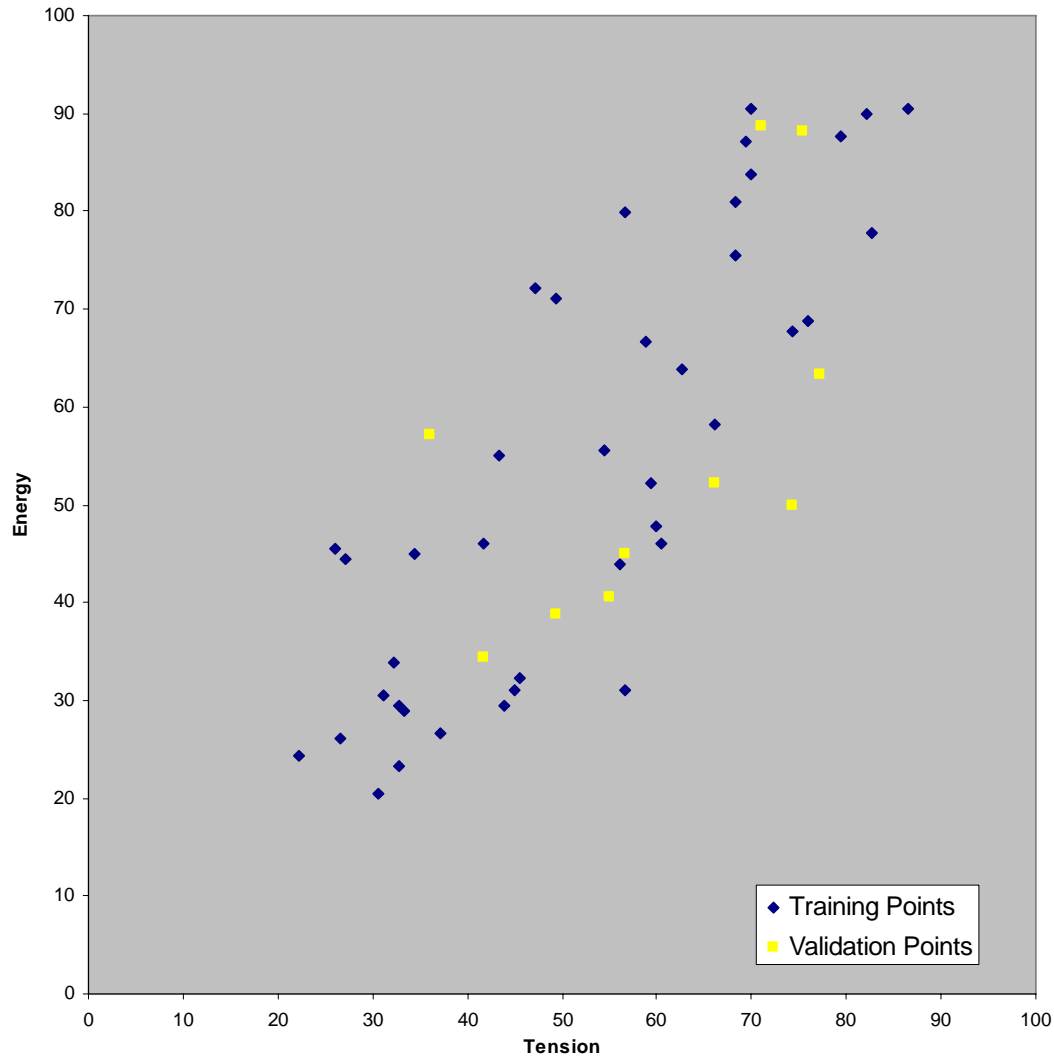
Machine Learning

- Used linear regression algorithm based on perceptrons (i.e. a neural network)
 - Permits real-valued outputs vs. classification
 - Permits multiple outputs
- Automatically fits a hypersurface to training data
- Query by finding points on the hypersurface
- Must avoid overtraining

Evaluation – human side

- Six survey participants
- Human results quite variable
- Distance from centroid most useful measure
 - Averaged of all participants for each clip
 - Averaged these averages
- NN validates based % error of energy and tension
 - Averaged similarly to above

Clip Mood Centroids from Human Evaluation



Evaluation – NN – 1 Hidden Layer

- 4 and 12 nodes tested
- Average error reduced to $< 1\%$ quickly (75-375 cycles)
- Would not validate test sets within 50% accuracy
- 4 and 12 node results comparable, but 12 node trained faster
- Cross-validation results were reasonably consistent

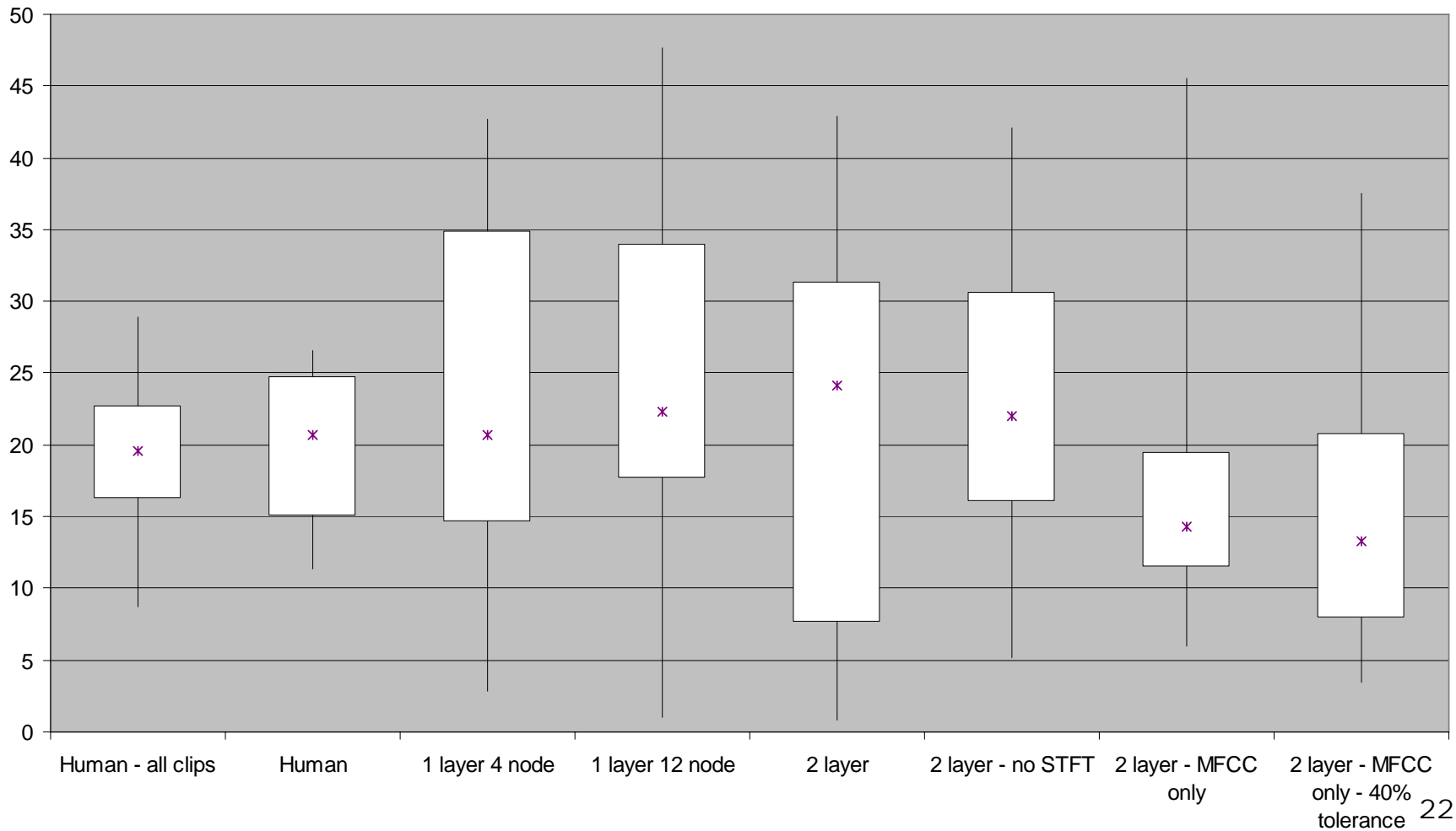
NN – 2 Hidden Layers

- Average error reduced to $< 1\%$ (270 cycles)
- Validated test sets within 50% accuracy for all samples quickly (400 learning cycles)
- Cross-validation results were reasonably consistent

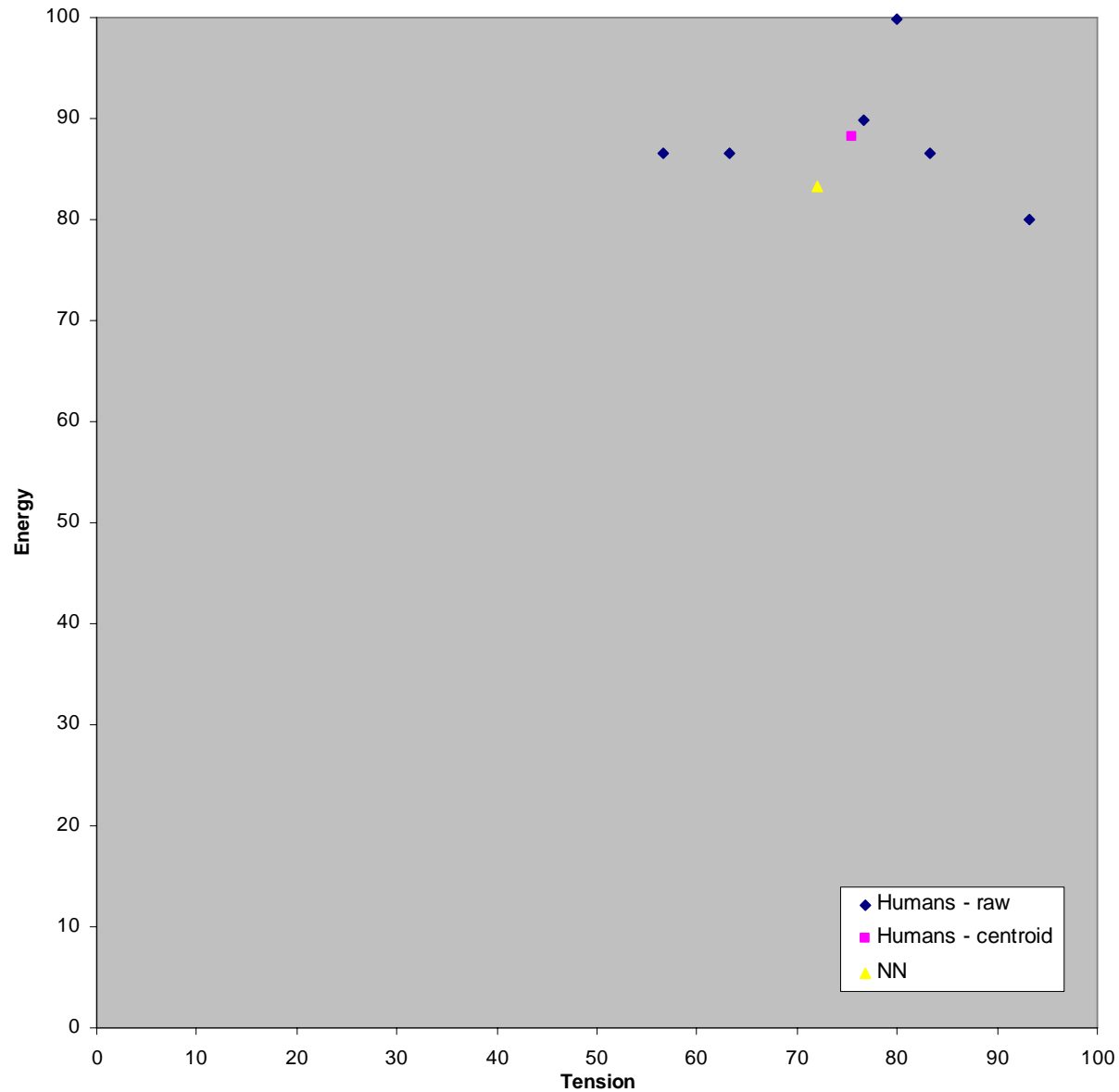
Feature analysis

- Repeated 2 layer tests with features removed
 - STFT+ZC removed
 - MFCC removed
 - BH removed
 - STFT+ZC and BH removed (only MFCC remaining)
- With STFT+ZC removed, results comparable
- With MFCC or BH removed, system would not validate
- With just MFCC, results were **better**

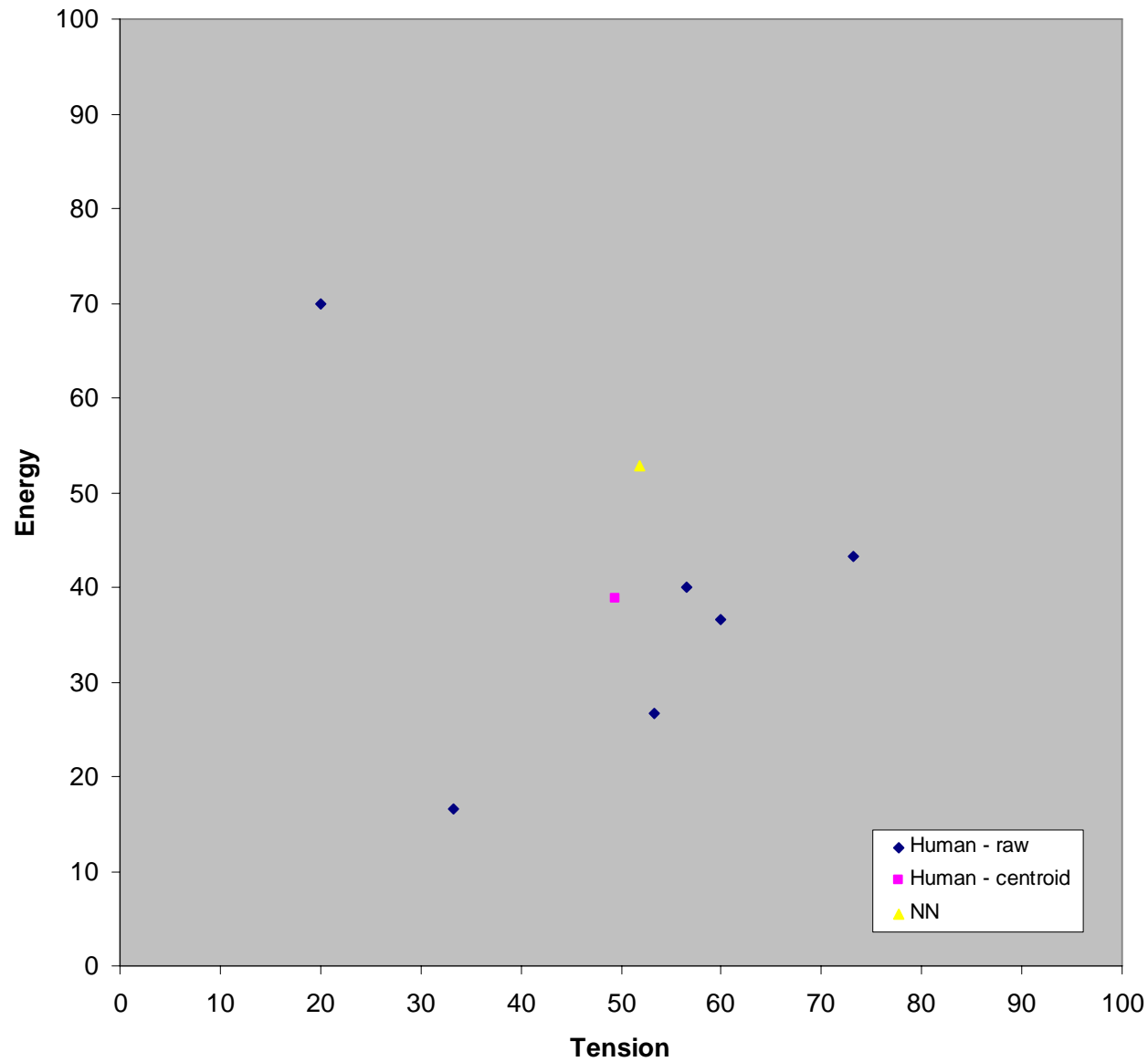
Comparative Results of Mood Quantification Methods



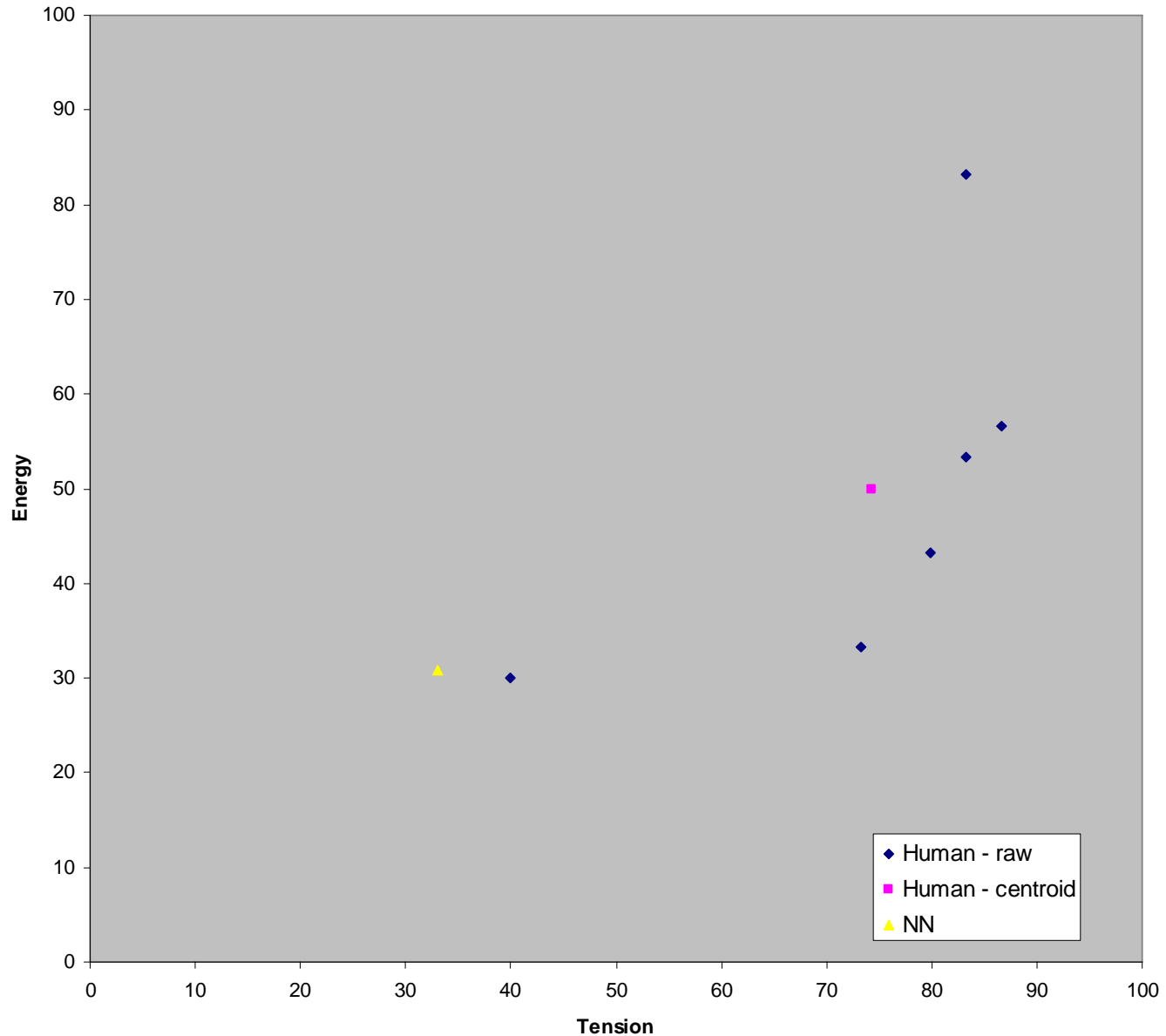
Best NN Performance



Median NN Performance



Worst NN Performance



Discussion

- Results promising – further research warranted
- Automated quantification **was** comparable to average human quantification
- Human quantification so variable that this is not so difficult a feat
- Refinements may improve performance on both human and machine sides

Possible Refinements

■ Human side

- Eliminate outlier human per clip before taking centroid
- Simpler human survey

■ Machine side

- Better selection of validation points to avoid edges

■ Overall

- More diverse clip selection

Future research

■ Pitch histogram features

- Based on specific pitches and relative pitches in song
- May help detect mood based on chord progressions and dissonance

■ Expand mood measure across entire song by segmentation

■ Study placebo effect

Key References

- Liu, D., Lu, L., & Zhang, H. (2003) Automatic Mood Detection from Acoustic Music Data. *Proc. Int. Symp. Music Information Retrieval (ISMIR) 2003*.
- Sarle, W. (ed.) (2002). *Comp.ai.neural-nets frequently asked questions*. Retrieved June 15, 2005 from <ftp://ftp.sas.com/pub/neural/FAQ.html>.
- Thayer, R. (1989). *The biopsychology of mood and arousal*. Oxford University Press.
- Tzanetakis, G. (2002). Manipulation, Analysis, and Retrieval Systems for Audio Signals, Ph.D. Dissertation, Princeton University



Questions?