



ON THE INFLUENCE OF QUANTIZATION ON THE IDENTIFIABILITY OF EMOTIONS FROM VOICE CODING PARAMETERS

Patrick Robitaille, Samuel Trempe, Philippe Gournay and Roch Lefebvre
Philippe.Gournay@USherbrooke.ca

Université de Sherbrooke
 Faculté de Génie
 2500, boul. de l'Université
 Sherbrooke (Québec)
 J1K 2R1 Canada

1. Summary

- Emotions play a major role in voice communication
- The quality of their reproduction by low bit rate voice coders has never been investigated so far
- Automatic identification is used to evaluate the influence of quantization on the identifiability of emotions from conventional voice coding parameters
- Results open new perspectives for speech coding

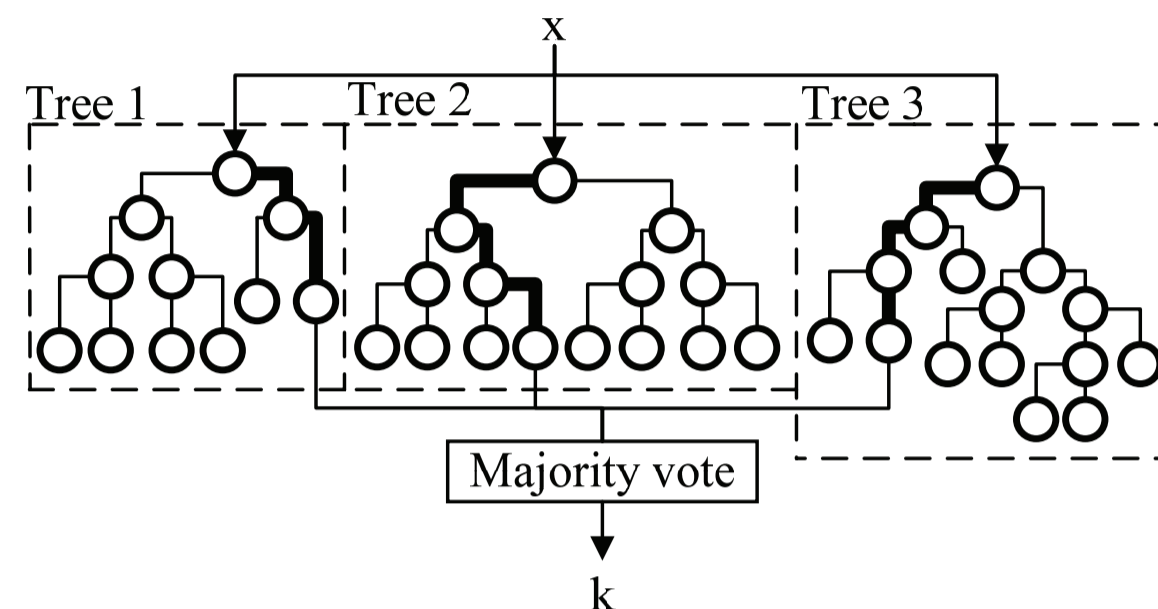
2. Feature Extraction

- Harmonic-Stochastic eXcitation analysis/synthesis model
- Typical voice coding parameter extraction
- Also, typical primary features for emotion identification:
 Pitch lag, voicing cut-off frequency, energy, LPC filter
- 8kHz sampling rate, 22.5ms frames

- 94 classification features derived from primary parameters

Primary	Secondary	Statistics
Pitch lag, Energy	First and second derivatives	Minimum, Maximum, Average, Range, Standard deviation, Kurtosis
	Jitter (pitch lag), Shimmer (energy)	
Voicing	Average length of voice segments, Percentage of voiced frames	
LPC coefficients	First and second formants, First and second derivative of those, Spectral slope coefficient	Minimum, Maximum, Average, Range, Standard deviation, Kurtosis

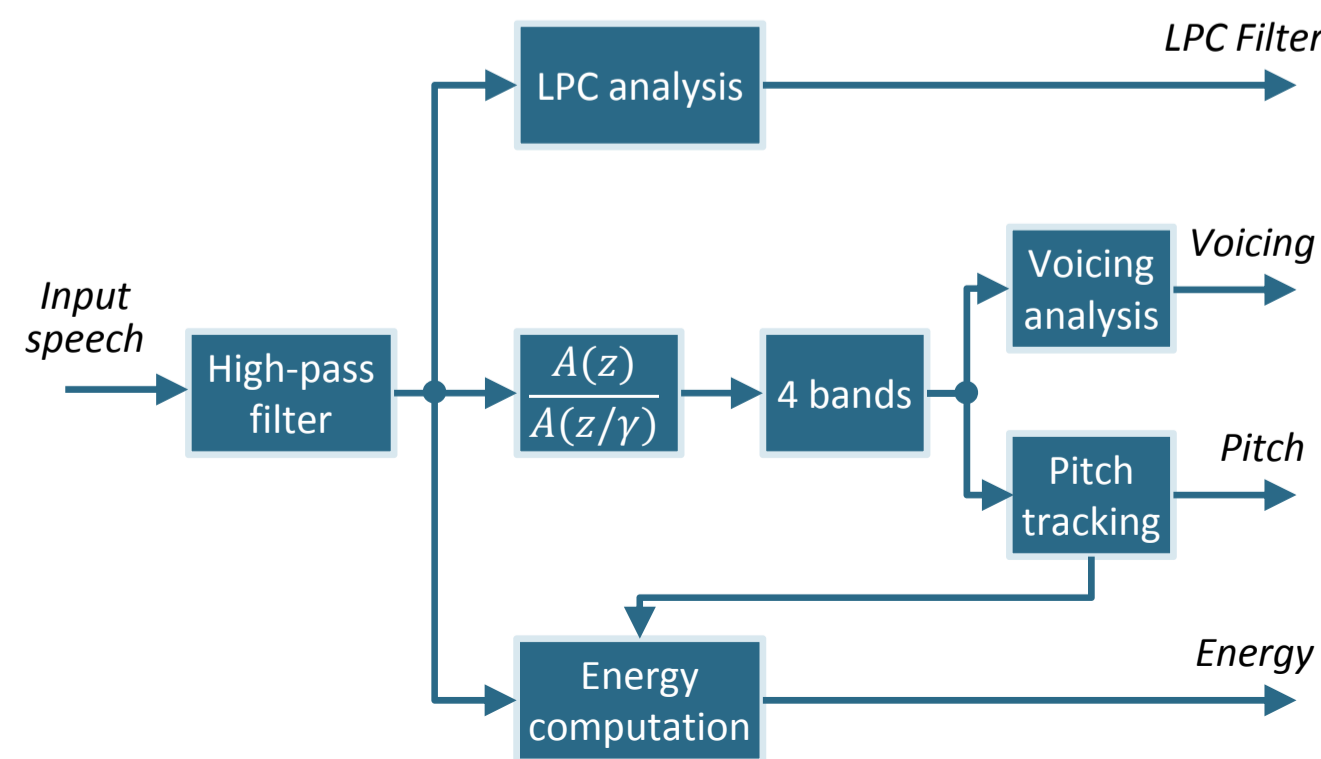
3. Random Forest Classifier



- Our typical random forest: 70 trees, 30 features per tree

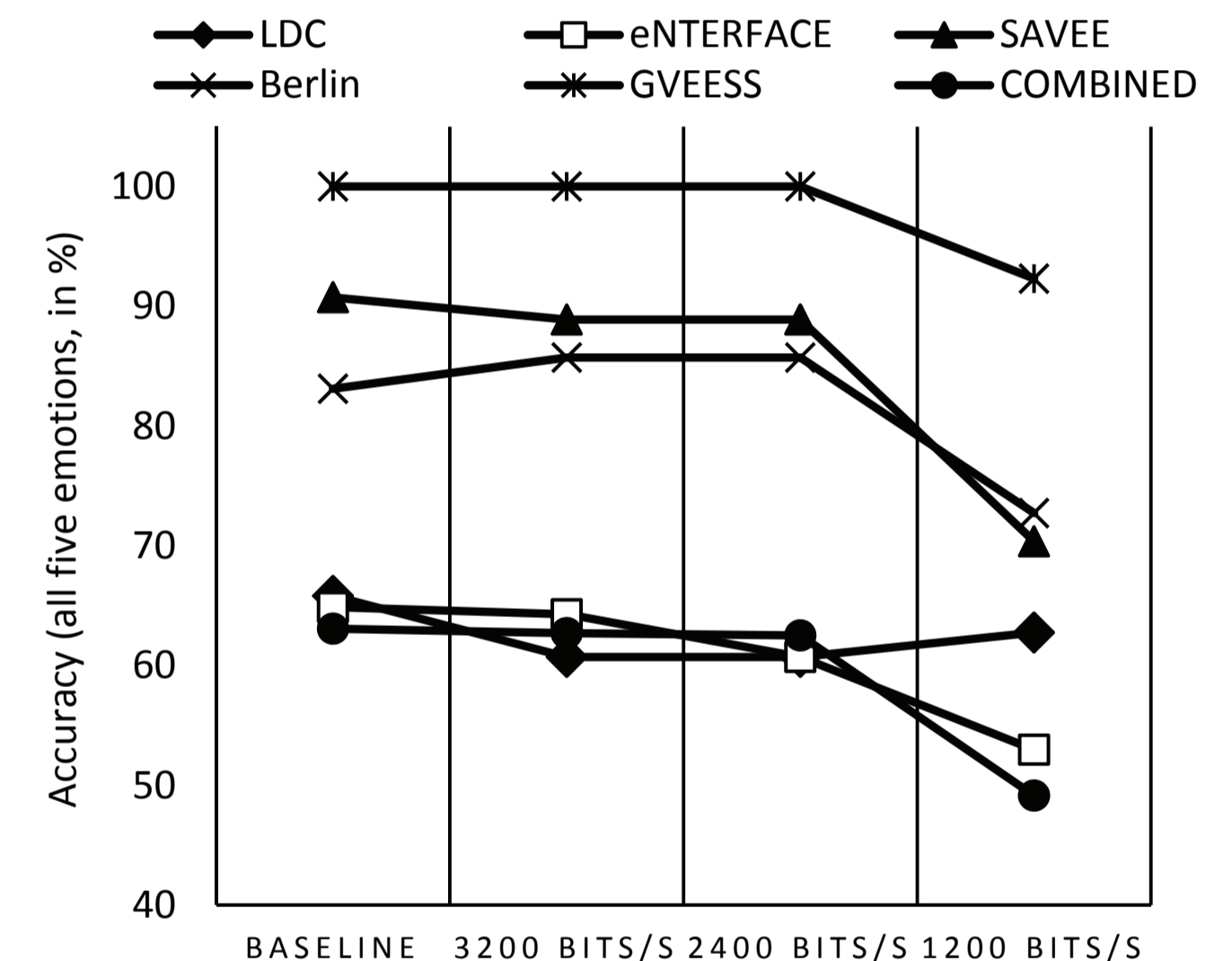
- Confusion matrix at 1200 bits/s (combined database)

Real \ Estim.	Anger	Sadness	Joy	Fear	Neutral
Anger	49.18 %	13.93 %	18.03 %	12.3 %	6.56 %
Sadness	11.54 %	52.31 %	14.62 %	10 %	11.54 %
Joy	16.67 %	8.33 %	56.25 %	16.66 %	2.08 %
Fear	17.76 %	15.89 %	24.3 %	34.58 %	7.48 %
Neutral	6.58 %	15.79 %	14.47 %	7.89 %	55.26 %



4. Performance Evaluation

- Five emotional speech databases: Berlin, SAVEE, eINTERFACE, GVEESS and LDC Emotional Prosody Speech and Transcripts
- Four conditions of quantization for primary parameters: baseline (unquantized), 3200, 2400 and 1200 bits/s



5. Conclusions and Perspectives

- The emotional state of a speaker can be identified automatically using voice coding parameters
- The performance of this identification decreases when the parameters are quantized, especially at lower rates
- Quantization of speech coding parameters could be improved by targeting the faithful reproduction of higher-level voice communication attributes such as emotions, rather than optimizing objective measures such as signal-to-noise ratios, mean squared errors and spectral distortion