Robustness and independence of voice timbre features under live performance acoustic degradations

Dan Stowell and Mark Plumbley
dan.stowell@elec.qmul.ac.uk

Centre for Digital Music
Queen Mary, University of London

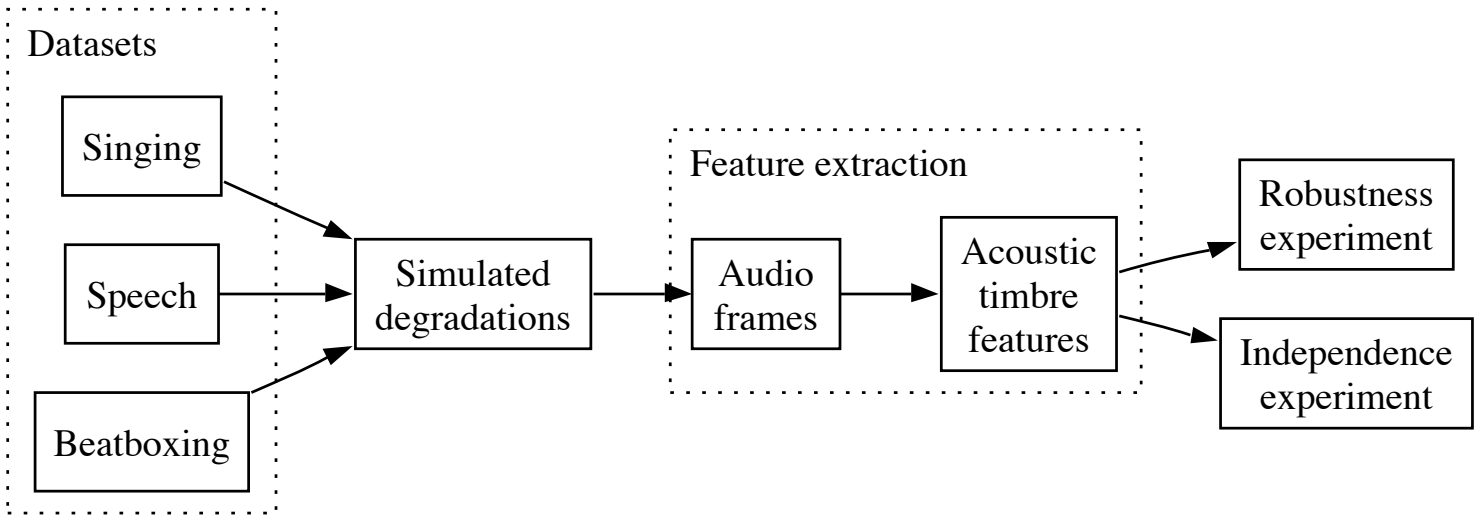
September 2008

centre for digital music
Queen Mary University of London

Introduction: Motivation

- Analysing timbre of performing voice
  - Create a timbre space
  - Input to classifier
  - Control effects
- Many acoustic features available
  - Cannot use all at once
- Desire those which
  1. Are most robust against noise/echo/etc
  2. Give us the most "information"
- Two experiments on *continuous-valued* features

Introduction: Two experiments



Features investigated

23 acoustic timbre features:

- MFCCs
- Spectral centroid
- Spectral spread
- Spectral crest factors (overall and subband)
- Spectral percentiles: 25%, 50%, 90%, 95%
- High-frequency content (HFC)
- Zero-crossing rate (ZCR)
- Spectral flatness
- Spectral flux

Robustness: method

- 7 types of degradation:
  - White noise
  - Crowd noise
  - Music noise
  - Clipping distortion
  - Delay
  - Delay with feedback
  - Reverb

(Each at 4 effect levels)

Measure absolute % deviation within each frame.

Two ways of comparing:

- Ranking
  (+ Kendall's W test)
- Pairwise comparison
  (+ Wilcoxon Signed-Rank test)

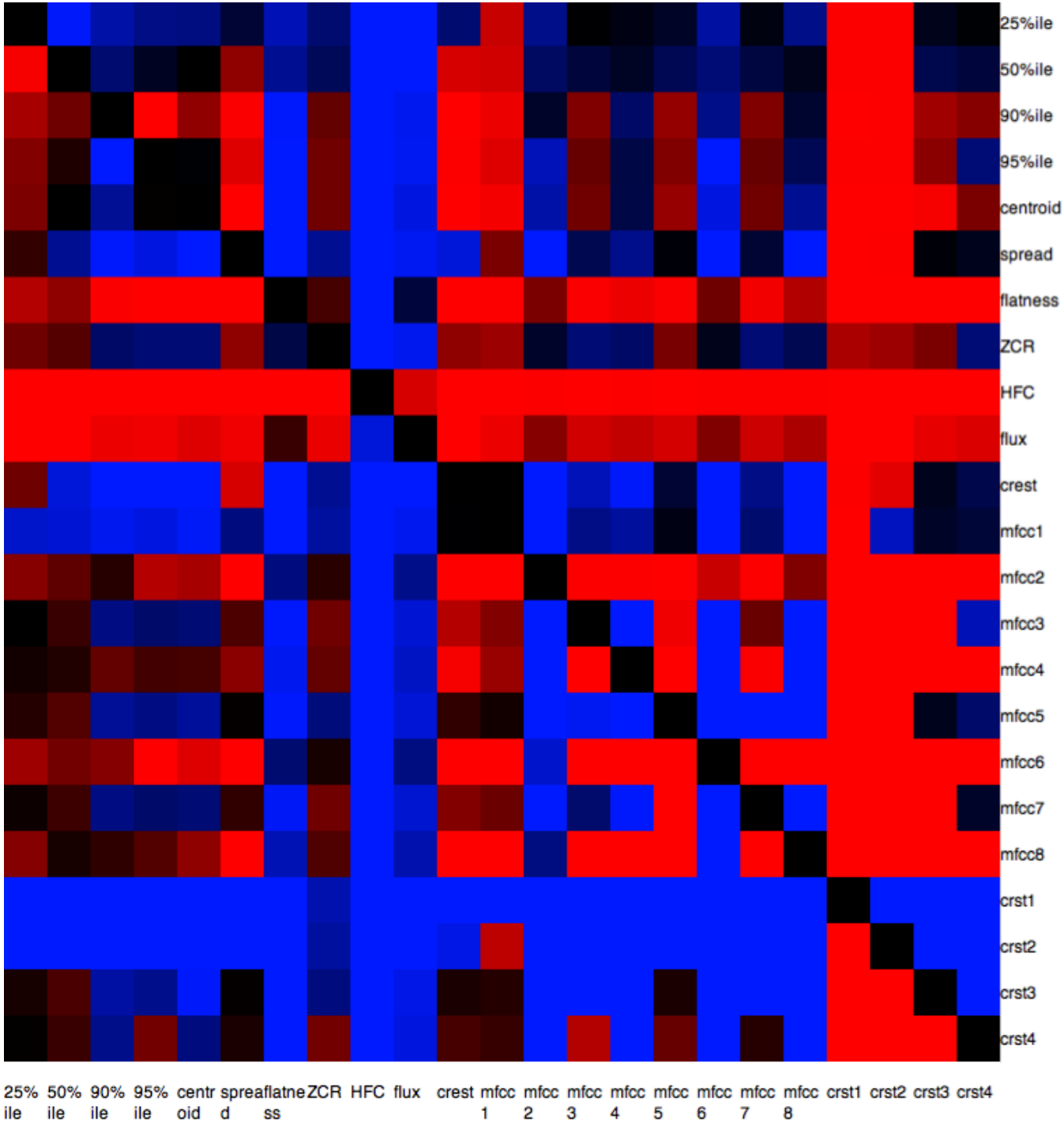
Robustness: results

Dataset	Singing	Speech	Beatboxing
BEST	crst1	crst1	crst1
	25%ile	mfcc1	mfcc5
	crst2	crst2	mfcc7
	ZCR	25%ile	mfcc1
	mfcc1	spread	mfcc3
	95%ile	crest	crest
	spread	50%ile	mfcc8
	crest	mfcc5	spread
	50%ile	crst3	mfcc6
	crst3	ZCR	mfcc4
	90%ile	mfcc7	25%ile
	centroid	mfcc3	crst2
	...	...	...

Robustness: results

Dataset	Singing	Speech	Beatboxing
	...	...	...
	centroid	mfcc3	crst2
	mfcc3	95%ile	crst3
	crst4	centroid	50%ile
	mfcc5	crst4	95%ile
	mfcc8	90%ile	crst4
	mfcc7	mfcc4	centroid
	flatness	mfcc8	90%ile
	mfcc4	mfcc2	ZCR
	mfcc2	mfcc6	mfcc2
	flux	flatness	flatness
	mfcc6	flux	flux
WORST	HFC	HFC	HFC

centre for digital music | Queen Mary University of London


25%ile
50%ile
90%ile
95%ile
centroid
spread
flatness
ZCR
HFC
flux
crest
mfcc1
mfcc2
mfcc3
mfcc4
mfcc5
mfcc6
mfcc7
mfcc8
crst1
crst2
crst3
crst4
25% ile
50% ile
90% ile
95% ile
centr oid
sprea d
flatne ss
ZCR
HFC
flux
crest
mfcc 1
mfcc 2
mfcc 3
mfcc 4
mfcc 5
mfcc 6
mfcc 7
mfcc 8
crst1
crst2
crst3
crst4

Robustness: results

- Some good:
  - Spectral crest factors
  - Odd-numbered MFCCs
- Some poor:
  - HFC
  - Spectral flatness
  - Spectral flux
  - Some even-numbered MFCCs
- Some interact with signal type:
  - ZCR
  - Some spectral percentiles

2. Independence

- Second experiment:
  Which features "give us the most information"?
- There may be redundancy between acoustic features
  - Correlation is one way to probe this – but limited (monotonic)
- Information theory: analyse dependencies more generally
- Again, two comparisons:
  - Pairwise
  - Ranking (feature selection)

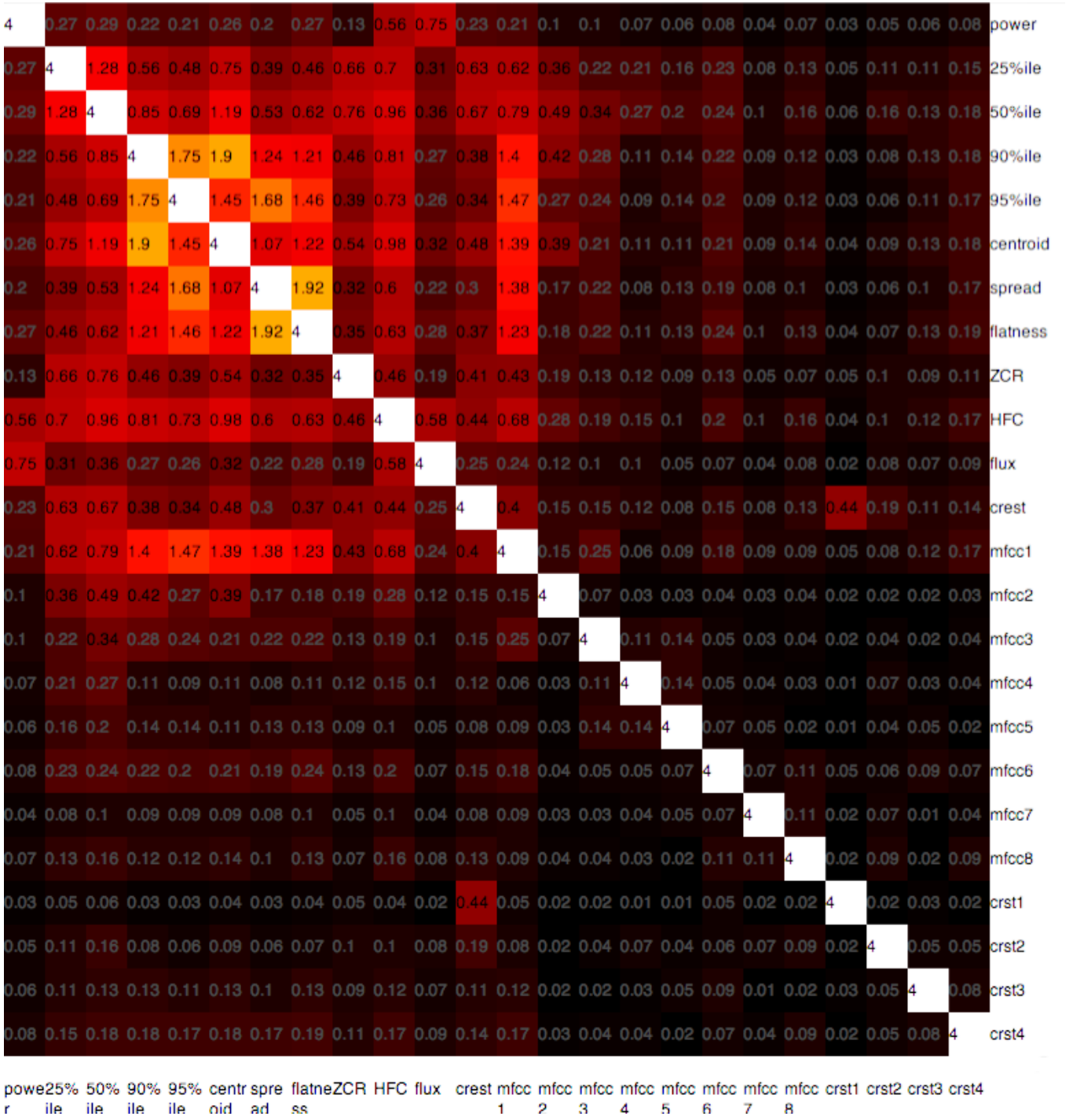
Independence: method (a)

Mutual information:

- Given feature X and feature Y:
  - If I know the value of X, how far does that decrease my uncertainty about the value of Y?
- Defined from the probability distributions:

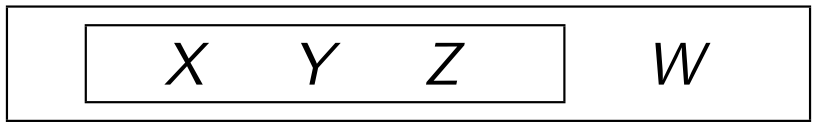
$$I(X;Y) = \sum_{y \in Y} \sum_{x \in X} p(x,y) \log \left(\frac{p(x,y)}{p(x)\,p(y)} \right)$$

- We can estimate this value from our data
- Tell us which features have informational overlap


power
25%ile
50%ile
90%ile
95%ile
centroid
spread
flatness
ZCR
HFC
flux
crest
mfcc1
mfcc2
mfcc3
mfcc4
mfcc5
mfcc6
mfcc7
mfcc8
crst1
crst2
crst3
crst4

Independence: method (b)

Conditional entropy:



Entropy of W *conditional on* X, Y, Z

$$H(W|X,Y,Z) = H(X,Y,Z,W) - H(X,Y,Z)$$
$$\not\equiv H(W)$$

Feature selection by greedy rejection:
reject one feature at a time,
according to lowest conditional entropy

Independence: results

	Singing	Speech	Beatboxing
BEST	crst2	crst2	crst1
	crst3	95%ile	mfcc1
	crest	crst1	crst2
	mfcc6	crst3	mfcc5
	mfcc8	mfcc8	mfcc7
	mfcc3	mfcc3	mfcc3
	crst1	mfcc7	mfcc8
	mfcc7	mfcc6	mfcc4
	95%ile	mfcc4	mfcc6
	mfcc4	mfcc5	crest
	mfcc5	crest	spread
	mfcc1	mfcc1	crst3
	spread	spread	95%ile
	90%ile	90%ile	crst4
	crst4	crst4	90%ile
	centroid	centroid	centroid
	ZCR	ZCR	ZCR
	50%ile	50%ile	50%ile
WORST	25%ile	25%ile	25%ile

centre for digital music

Queen Mary University of London

Summary

1. Robustness
   - Ranking (median deviation)
   - Pairwise comparison (Wilcoxon Signed Rank test)
2. Independence
   - Pairwise comparison (mutual information)
   - Feature selection (conditional entropy)

Conclusions

- Suggested feature-set for performing voice:
  - Spectral crest factors + MFCCs + 95-percentile *("rolloff")*
- Spectral crest factors perform well
- Spectral centroid less useful than expected
- Some features' performance interacts with signal type
- Information-theoretic measures useful for probing dependencies