Shaking and Speech-smile Vowels Classification: An Attempt at Amusement Arousal Estimation from Speech Signals

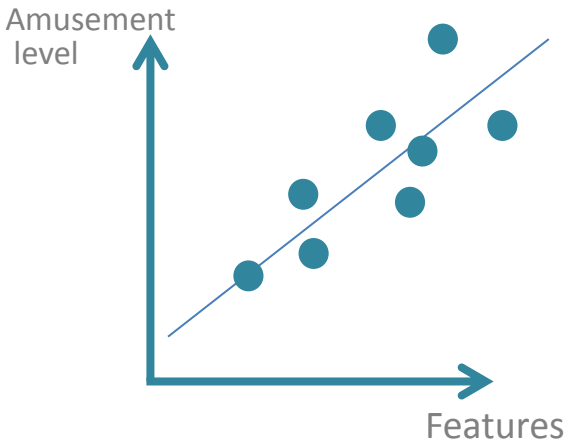
Kevin El Haddad, Stéphane Dupont, Hüseyin Cakmak, Thierry Dutoit

TCTS - University of Mons

Outlines

- Overview of Main Objective
- Approach
- Amusement Components Classification
- Conclusion
- Ongoing Work and Perspectives

Main project: Amusement Level Estimation



Amusement level assessment :

- Recognition of amusement component in speech
- Mapping between componants and amusement levels
- Contribution to context understanding
- Real-time system

JOKER

http://www.chistera.eu/projects/joker

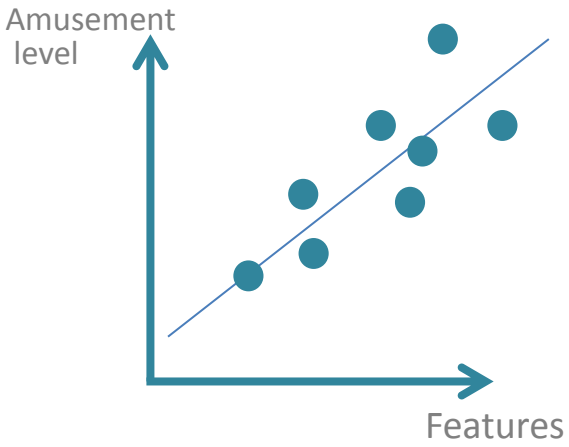
Purpose:

- Contribution in HCI
- User emotional state estimation on an amusement scale

Main project: Amusement Level Estimation

Amusement level assessment :

- Recognition of amusement component in speech
- Mapping between componants and amusement levels
- Contribution to context understanding
- Real-time system



Approach

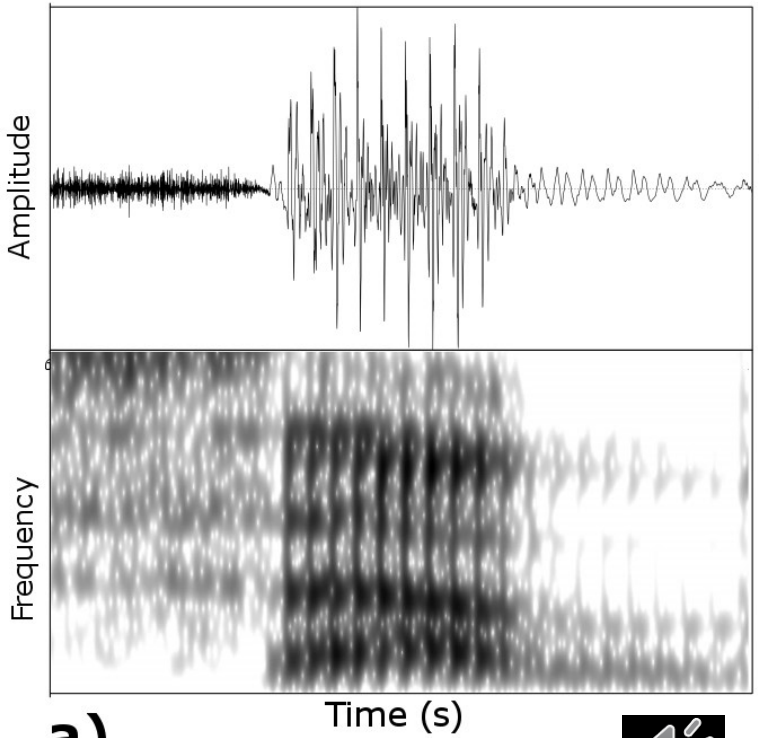
Usual approach:

- Extract global features (pitch, MFCC, etc..)
- Use them for Emotion classification and dimension estimation

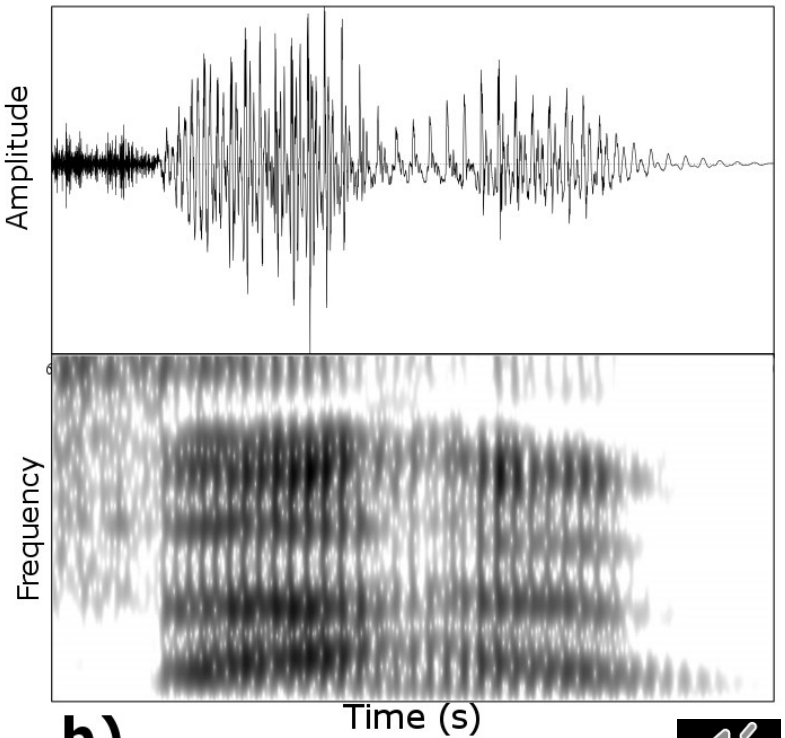
Our approach:

- Recognize amusement components in speech
- Two components are focused on in this work:
  1. Smiled vowels
  2. Shaking vowels
- Map detected components to arousal level for amusement

Amusement Components Classification: The components



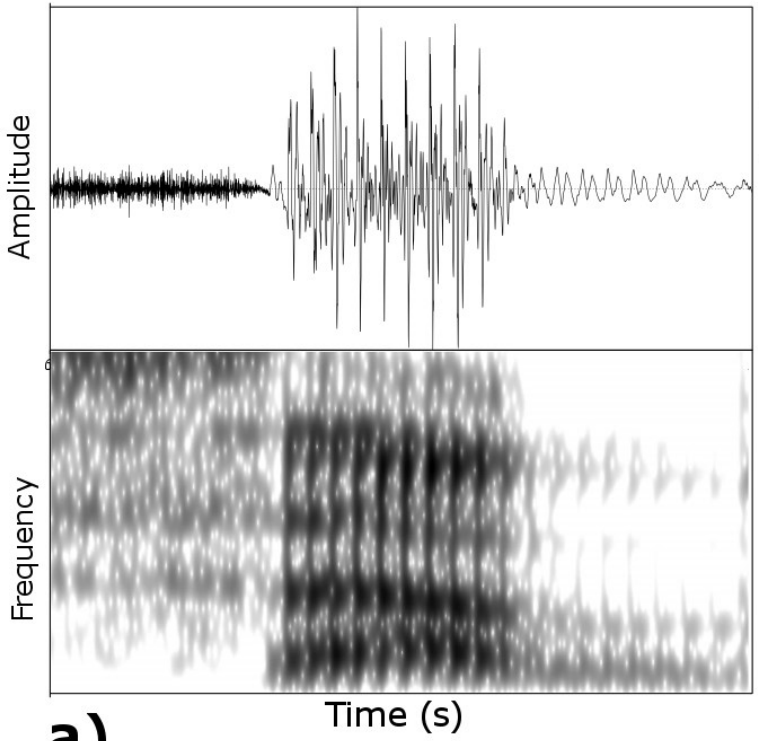
a) Smiled Vowel



b) Shaking Vowel

- **Vowel-like signals**
- **Discontinuity in spectral domain due to air burst**

Amusement Components Classification: The components

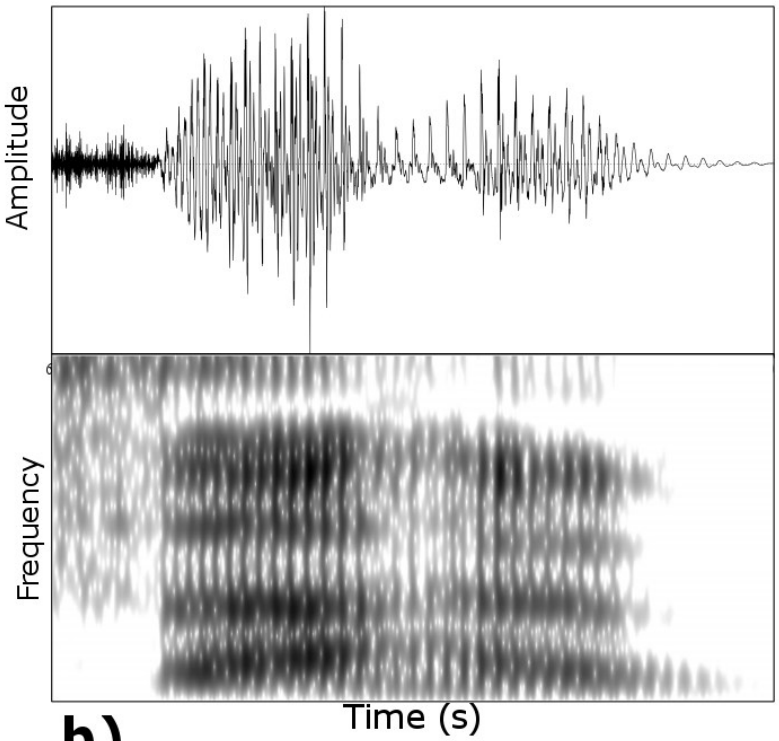


a) **Smiled Vowel**

335 samples

ICSI

TCTS data



b) **Shaking Vowel**

48 samples

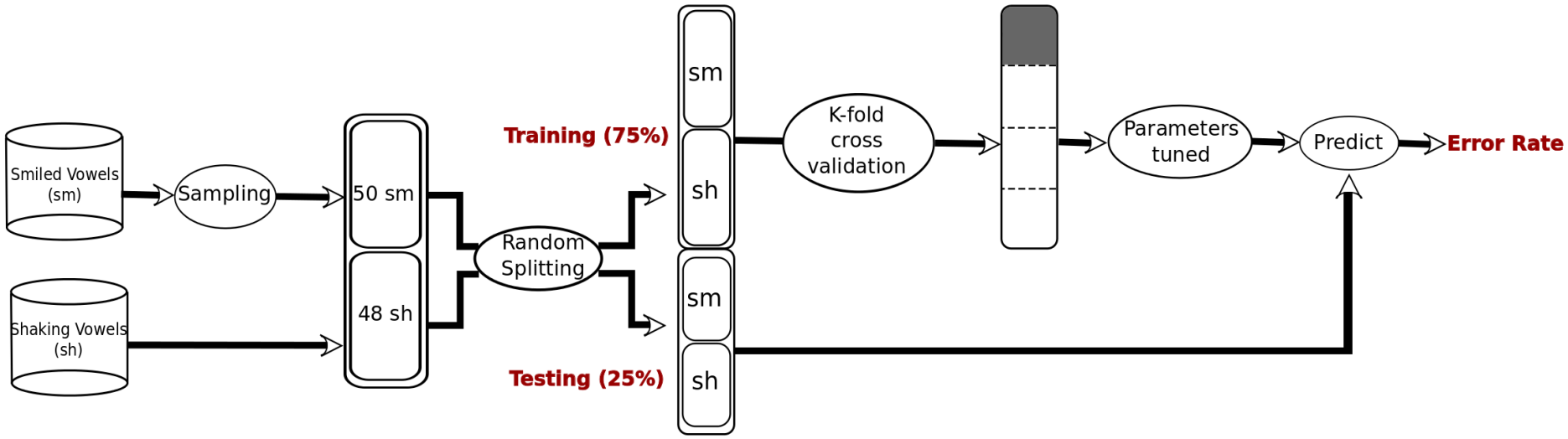
ICSI

TCTS data

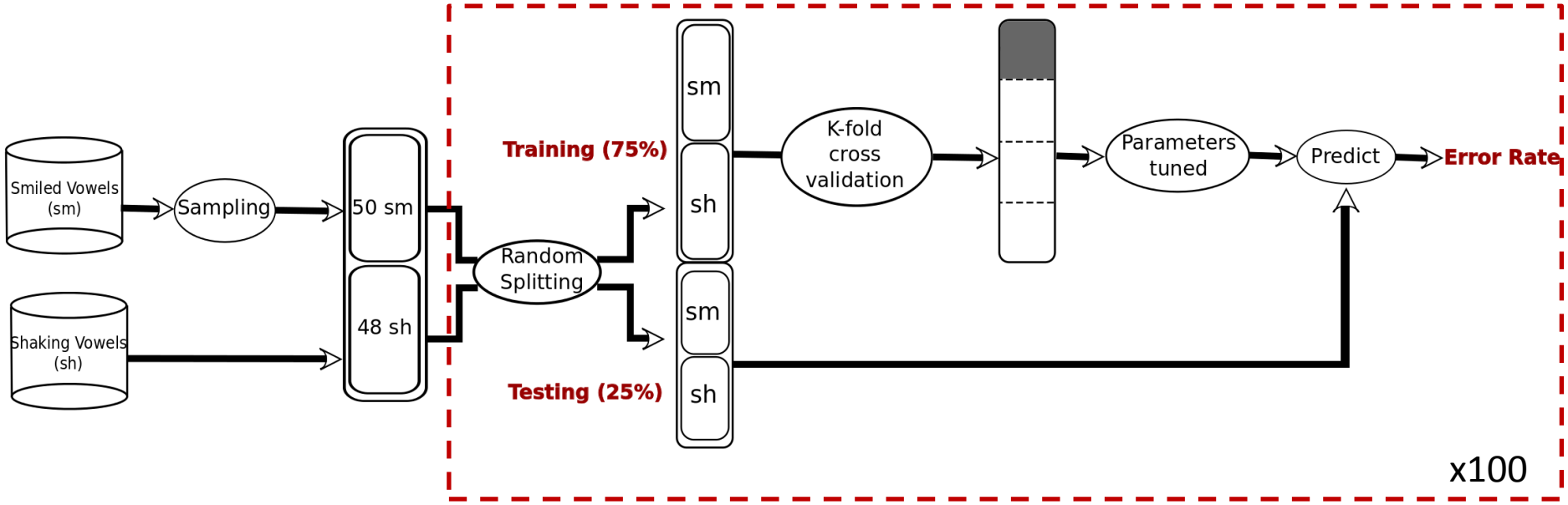
Amusement Components Classification: Feature extraction

	Set 1	Set 2	Set 3
Parameters	**Mean and StD of:** • 13 MFCC • 13 Δ MFCC	• Positive Negative Amp. Ratio • Spectral flatness **Mean and StD of:** • F0 • Spectral Centroid • Max. Voiced Freq • Energy	**StD of:** • Δ F0 and log-power envelope • Residuals of F0 and log-power envelope to linear regression
Description	Frequently used in speech recognition systems	Spectral and temporal features	Stability-Based features (**New!)**

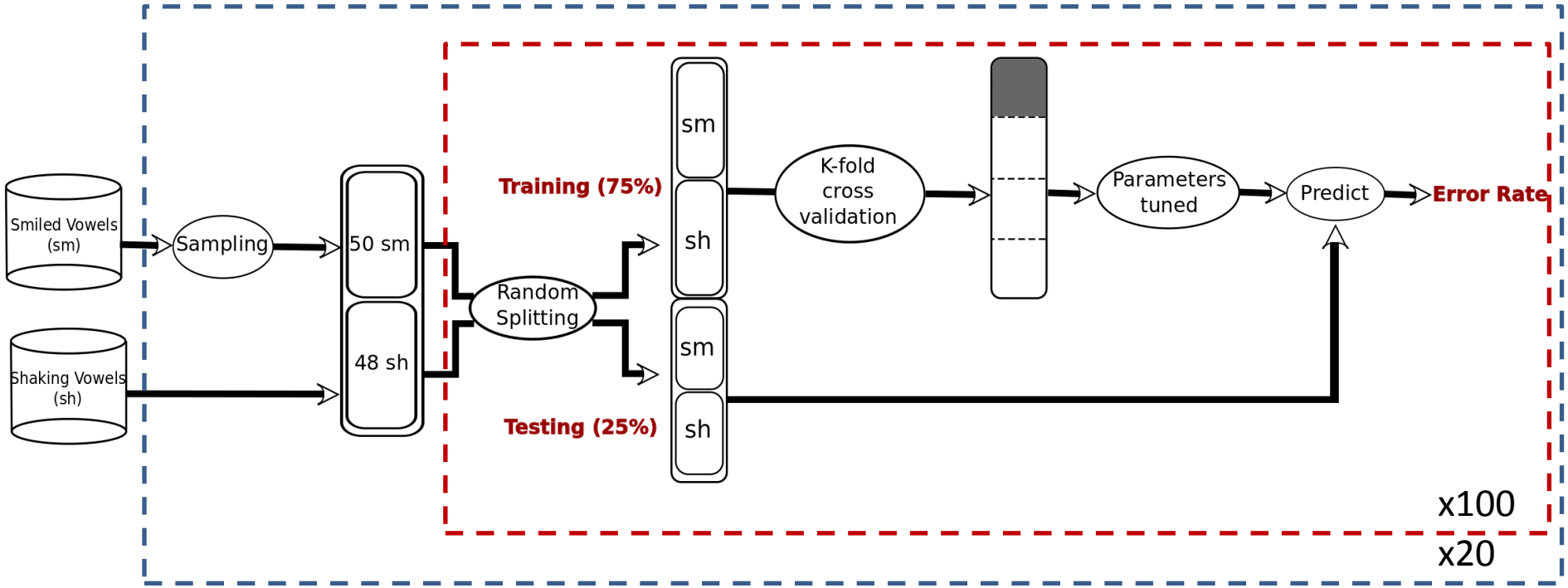
Amusement Components Classification: Machine Learning Approach



Amusement Components Classification: Machine Learning Approach



Amusement Components Classification: Machine Learning Approach



Amusement Components Classification: Machine Learning Approach

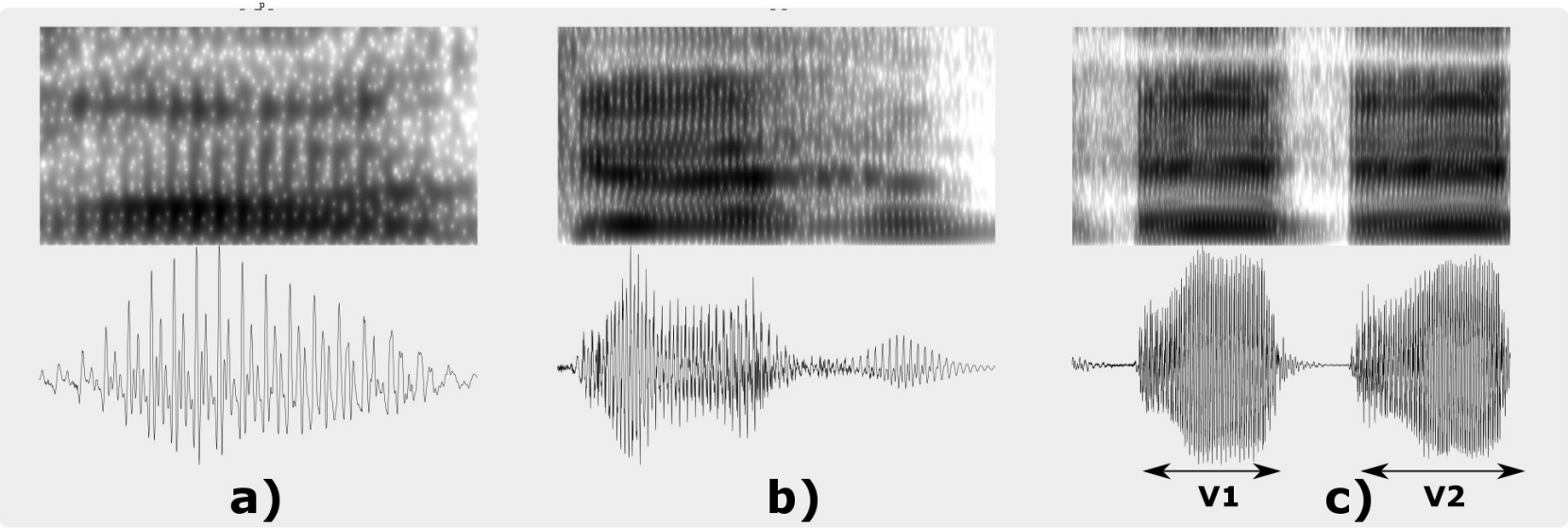
Error Rates of a k-Nearest Neighbor (kNN):

Sets	Mean	StD
Set 1	32%	2.5%
Set 2	33.6%	2.3%
Set 3	32.7%	3.2%
Set 1 + Set 2	32%	2.5%
Set 1 + Set 3	32.5%	3.6%
Set 2 + Set 3	**30.1%**	2.8%
All	32.1%	2.7%

Conclusion

- Database gathered
- Stability-Based Features (SF) introduced
- SF useful for smiled/shaking vowel classification (contribution to a lower error rate)

Ongoing and Future Work



Smiled Vowel	Shaking Vowel	Laughter bursts
335	48	1004
ICSI TCTS data	ICSI TCTS data	TCTS data

K. El Haddad, S. Dupont, H. Cakmak, T. Dutoit, “Towards a Level Assessment System of Amusement in Speech Signals: Amused Speech Components Classification”, International Symposium on Signal Processing and Information Technology (ISSPIT 2015), pp. 12-17, Abu Dhabi, UAE, 7-10 December

Ongoing and Future Work

Features:

- F0:
  1. 20 ms window shifted by 10 ms
  2. Mean and StD of F0

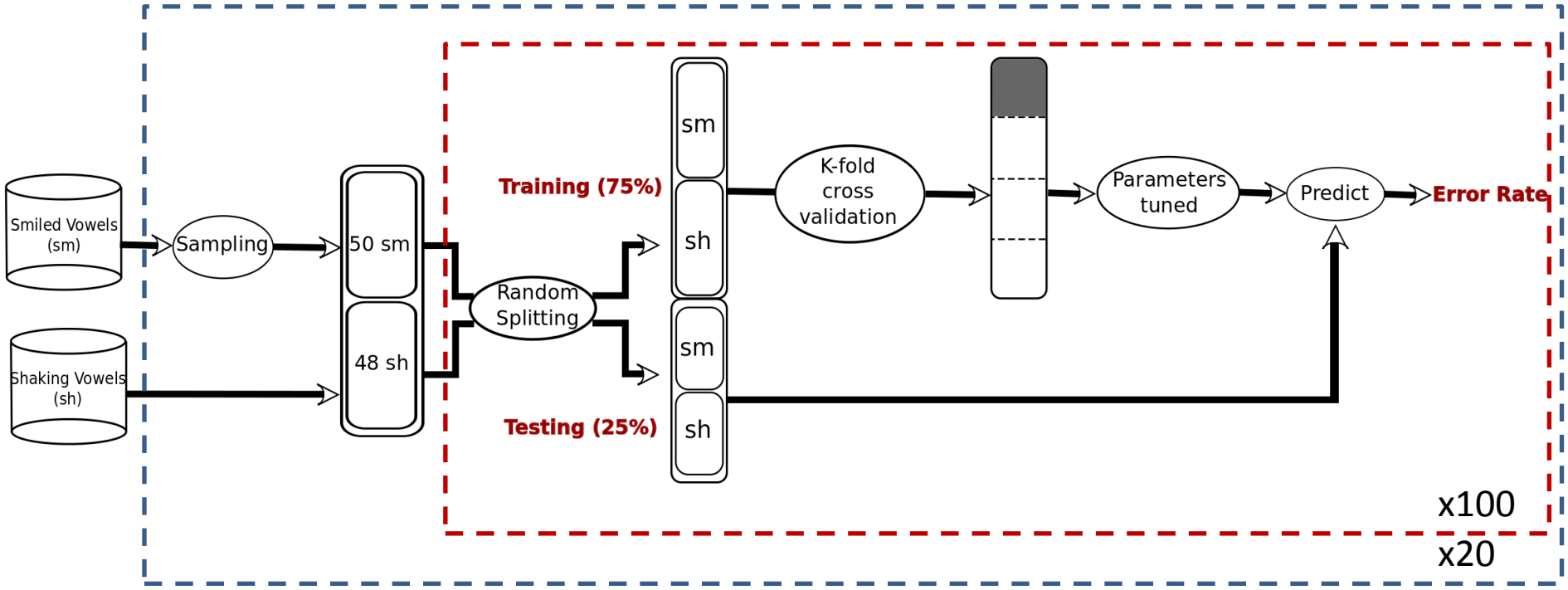
- MFCC:
  1. 20 ms window shifted by 10 ms
  2. Mean of each of 12 coefficients + 0th coefficients

- Stability-based features:
  1. StD of F0 and Δ F0 residuals
  2. StD of log Power residuals and Δ log Power
  3. Finally: log of 1) and 2) due to skewness

K. El Haddad, S. Dupont, H. Cakmak, T. Dutoit, "Towards a Level Assessment System of Amusement in Speech Signals: Amused Speech Components Classification", International Symposium on Signal Processing and Information Technology (ISSPIT 2015), pp. 12-17, Abu Dhabi, UAE, 7-10 December

Ongoing and Future Work

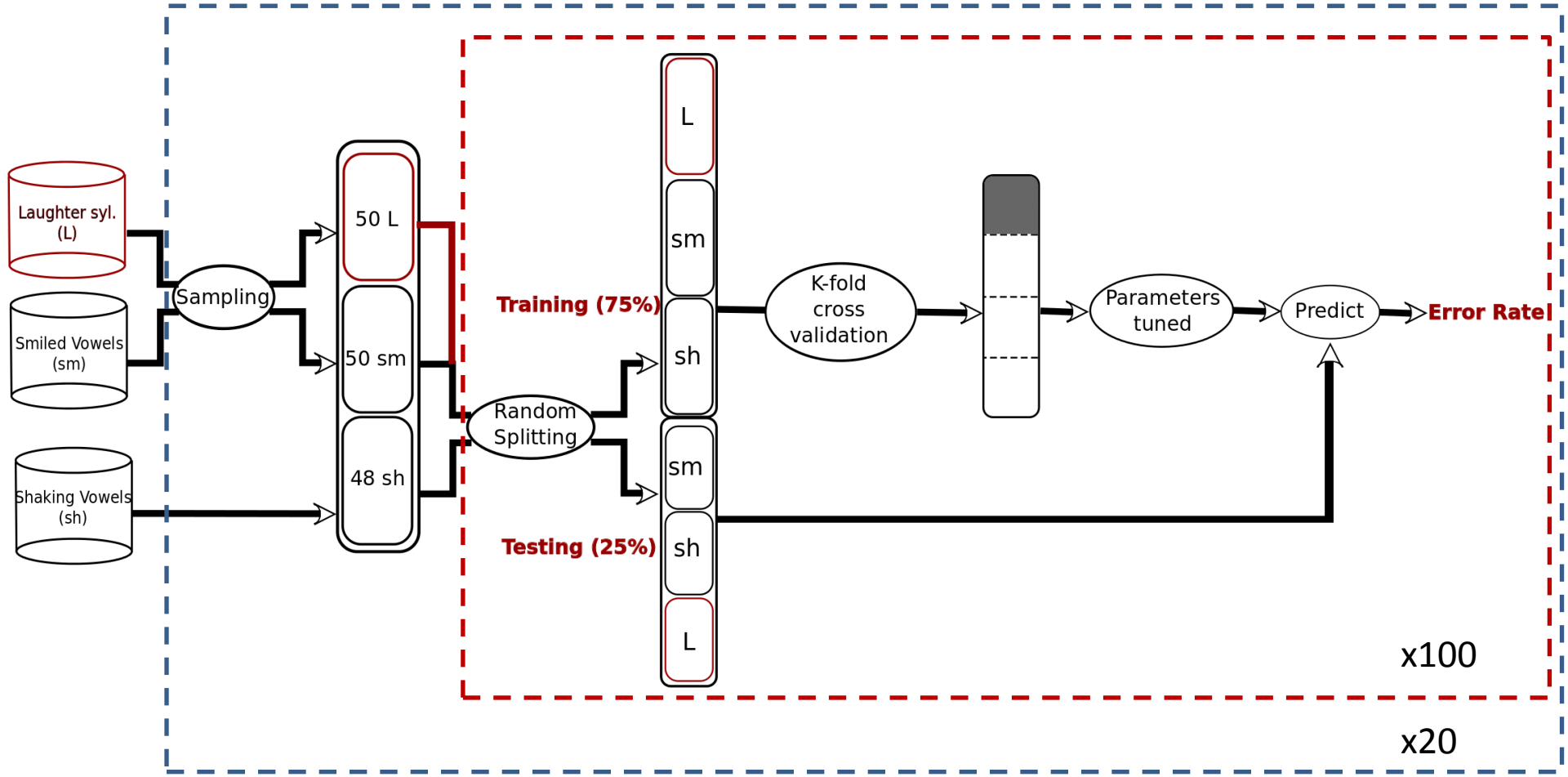
Previous pipeline:



K. El Haddad, S. Dupont, H. Cakmak, T. Dutoit, “Towards a Level Assessment System of Amusement in Speech Signals: Amused Speech Components Classification”, International Symposium on Signal Processing and Information Technology (ISSPIT 2015), pp. 12-17, Abu Dhabi, UAE, 7-10 December

Ongoing and Future Work

Recent pipeline:



K. El Haddad, S. Dupont, H. Cakmak, T. Dutoit, "Towards a Level Assessment System of Amusement in Speech Signals: Amused Speech Components Classification", International Symposium on Signal Processing and Information Technology (ISSPIT 2015), pp. 12-17, Abu Dhabi, UAE, 7-10 December

Ongoing and Future Work

Error Rates of different systems:

System	SF	MFCC	f0
kNN	28.2%	31%	33.9%
SVM-Lin	25.3%	31.2%	38.1%
SVM-Poly	24.4%	29.1%	30.8%
NN	**23.8%**	30.4%	29.9%

System	SF+MFCC	SF+F0	MFCC+f0	All
kNN	27.8%	25,9%	28.37%	26.9%
SVM-Lin	27.07%	23.1%	28.7%	25.2%
SVM-Poly	27.7%	**21.04%**	27.4%	27.6%
NN	26%	21.9%	27,9%	25%

K. El Haddad, S. Dupont, H. Cakmak, T. Dutoit, "Towards a Level Assessment System of Amusement in Speech Signals: Amused Speech Components Classification", International Symposium on Signal Processing and Information Technology (ISSPIT 2015), pp. 12-17, Abu Dhabi, UAE, 7-10 December

Perspectives

- Increasing the amount of data
- New features? New technique (Shapelets)?
- Mapping the components to amusement levels
- Real-time system