Fractal Nature of Chewing Sounds

Vasileios Papapanagiotou
Christos Diou
Anastasios Delopoulos

Multimedia Understanding Group
Dpt. Electrical & Computer Engineering
Aristotle University of Thessaloniki
Greece

Zhou Lingchuan

CSEM SA
Landquart
Switzerland

Janet van den Boer
Monica Mars

Division of Human Nutrition
Wageningen University
Netherlands

Dietary Monitoring

Using a non-intrusive, in-ear microphone

Audio recording throughout the entire day

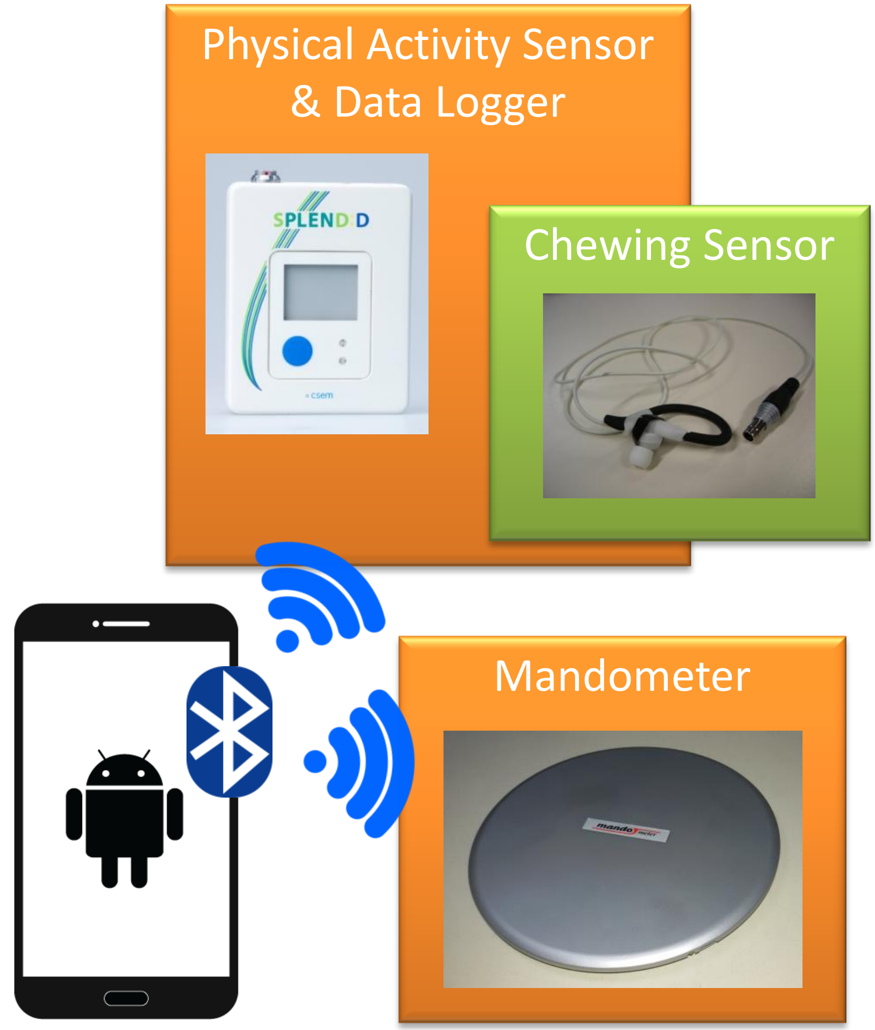
Detection of snacks for objective behavioural monitoring

The SPLENDID Project

SPLENDID

Personalised Guide for Eating and Activity Behaviour for the Prevention of Obesity and Eating Disorders

- Whole-day monitoring of dietary habits using chewing sensor
- Whole-day monitoring of physical activity level using physical activity sensor
- In-meal beahaviour monitoring using Mandometer
- Integration with an Android mobile phone over Bluetooth



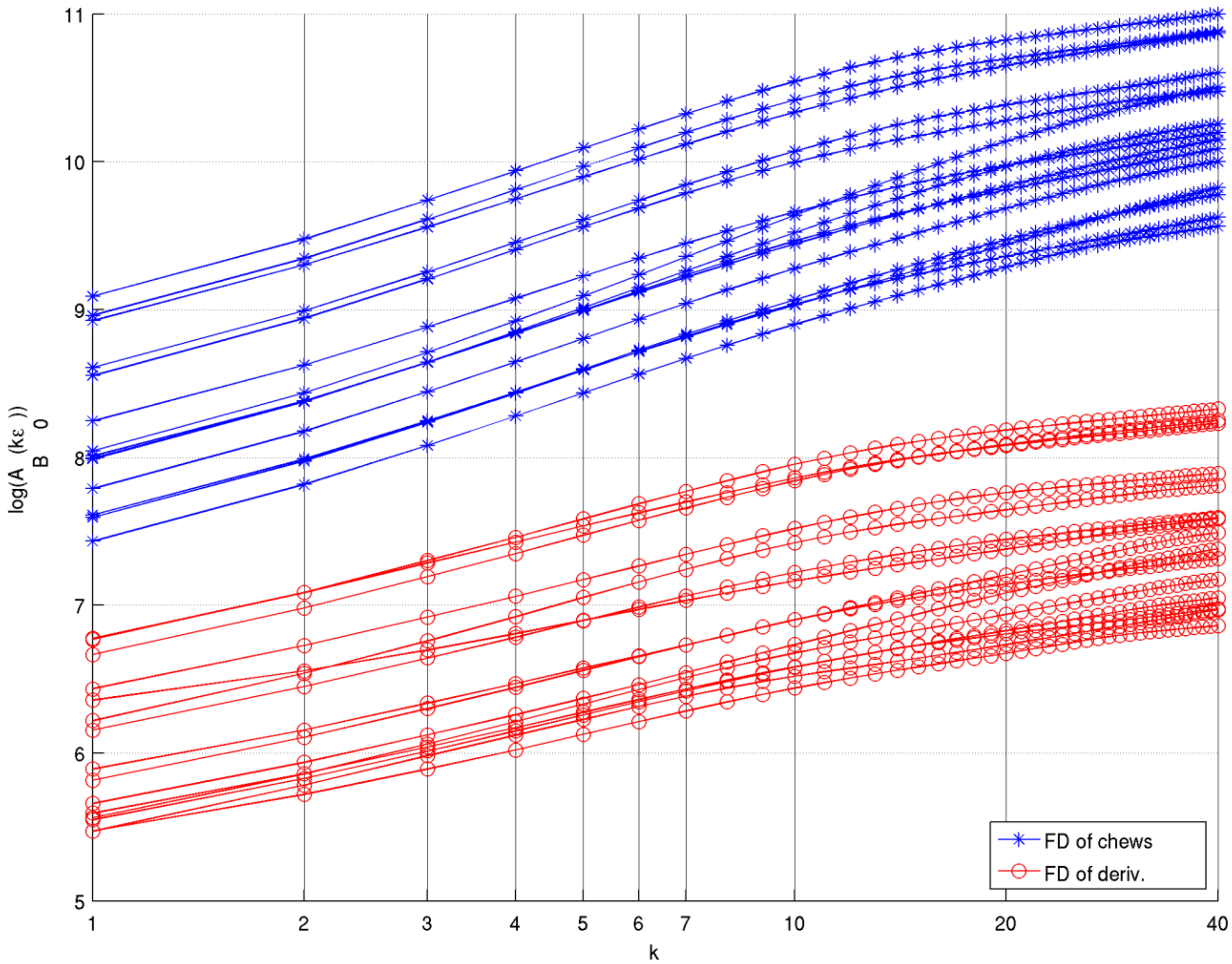
Motivation & Problem Statement

Main Problem	Detect chewing activity with an in-ear microphone and use it to create dietary behaviour reports
Challenges	Need for robust chewing detection
	No assumption about level of ambient noise
	Interference by talking, socializing, performing outdoor activities, etc
	Low computational effort, to enable mobile integration
Approach	Examine the Fractal Nature of chewing sounds, compared to other sounds
	Form a robust detector based on the Fractal Dimension of captured audio

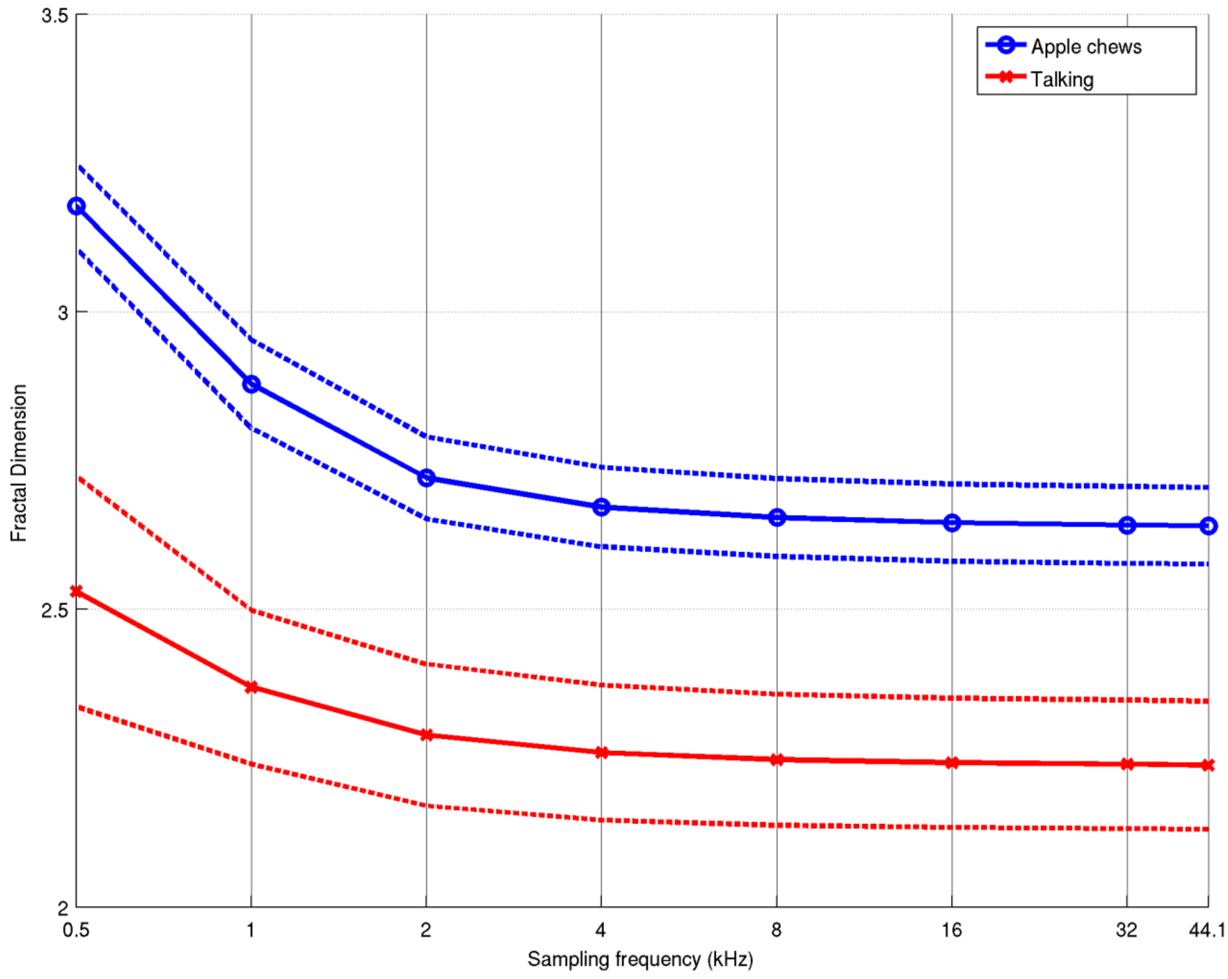
Fractal Dimension

Fractal Dimension by Mandelbrot	$D = 2 - \lim_{\varepsilon \to 0} \frac{\log A_B(\varepsilon)}{\log \varepsilon}$ For truly fractal signals, this ratio is independent of ε
The area $A_B(\varepsilon)$ is approximated using dilation and erosion banks, by	$A_B(\varepsilon) \approx \sum_{n=0}^{N-1} \left[x_k^d(n) - x_k^e(n)\right]$ $\varepsilon = k\varepsilon_0, \qquad k = 0, 1, 2, \cdots, M$
Data points	Linear fitting on the data points $(k, A_B(\varepsilon))$ can be used to compute the Fractal Dimension as the gradient of the fitted curve
The Fractal Dimension can thus be approximated using data points as	$D = \frac{1}{M} \sum_{k=1}^{\mathrm{M}} \frac{\log \mathrm{A_B}((k+1)\varepsilon_0) - \log \mathrm{A_B}(k\varepsilon_0)}{\log(k+1) - \log k}$

Chewing Sounds are Fractal-like



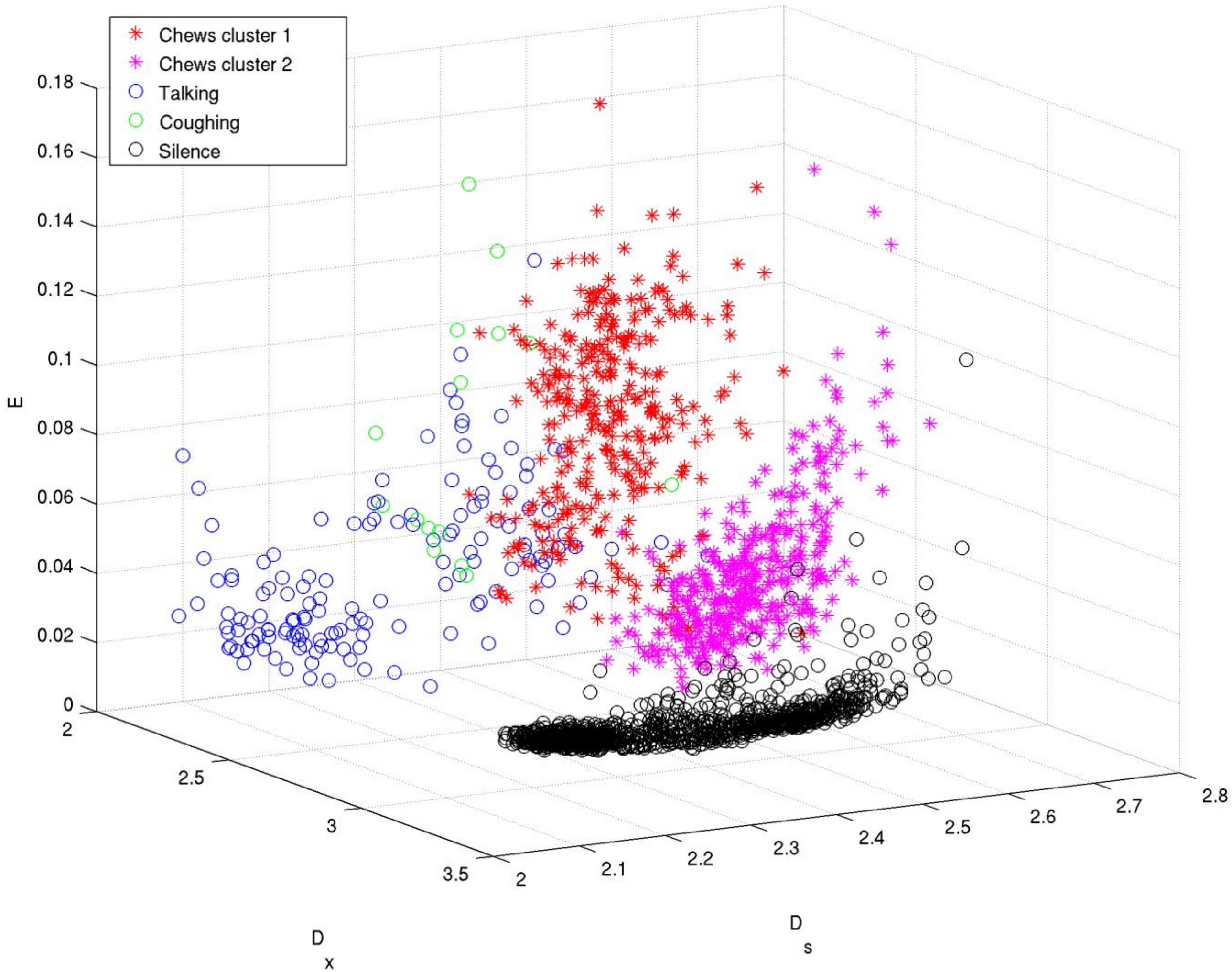
Higher Fractal Dim. of Chewing Sounds



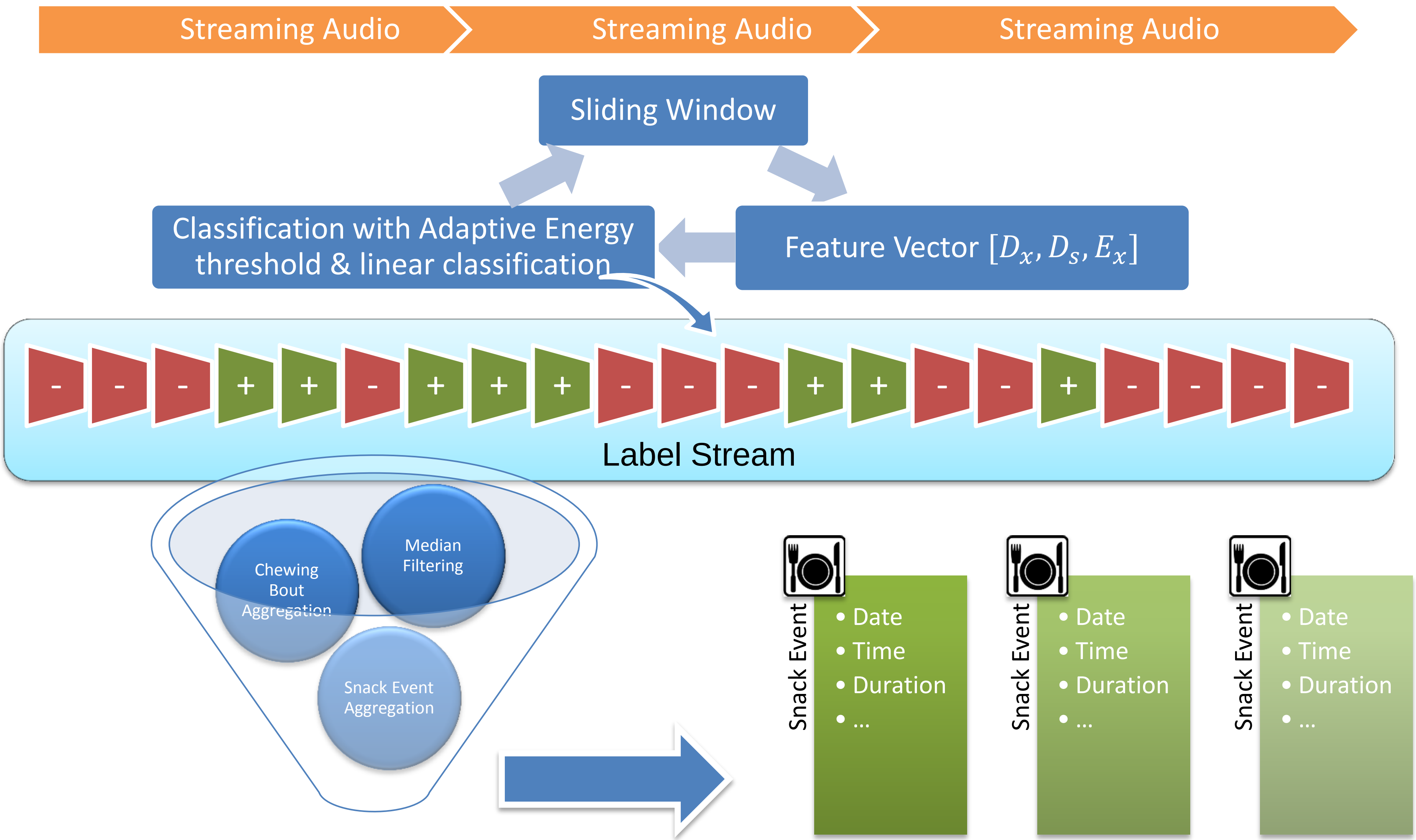
Designing a Detection Algorithm

Fractal Nature	Fractal dimension of the audio segments (and their derivatives) can be used to discriminate between chewing and non-chewing activities (such as talking)
	However, estimation of Fractal Dimension for low energy segments (silence or low ambient noise) is inaccurate
Features	D_x: Fractal Dimension of audio segment
	D_s: Fractal Dimension of derivative of audio segment
	E_s: Energy of audio segment
Decision	Energy threshold rejects silent/low-noise segments
	For non-rejected segments, a linear boundary discriminates between chewing and non-chewing

Feature Space



Detection Algorithm



Datasets

Complete Dataset

Collected at Wageningen University	10 subjects
	30 minute-long recordings
Various tasks	Eating various food types
	Talking
	Coughing
	Eating with Talking, background noise, etc
Prototype sensor by CSEM	FG-23329 mic housed in an ear bud

Extracted Chews

Food Type	No.	Type	No.
Apple	156	Cough	15
Banana	63	Pause	1032
Bread	84	Talk	147
Candy bar	96		
Chewing gum	126		
Potato chips	149		
Total	**674**		**1194**

Experimental Results for Extracted Chews

Parameters	Energy threshold: 0.02
	Boundary: $y = -2.62x + 8.73$
Training set accuracy	95.4 % for the three class task (chewing, talking/coughing, silence)
	96.5 % for chew vs. non-chew (chewing vs. talking/coughing)
Errors	Only 7 potato chip chews misclassified as talking/coughing
	Only 9 chews misclassified as silence

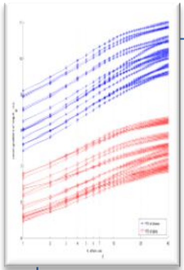
Class	Chew	T/C	Silence
Apple	**156**	0	0
Banana	**62**	0	3
Bread	**83**	0	1
Candy bar	**95**	0	1
Chewing gum	**120**	0	6
Cough	2	**13**	0
Pause	27	0	**1005**
Potato chips	**142**	7	0
Talking	21	**106**	20

Results for Chew Bouts and Snacks

Algorithm	Chew bout		Snack	
	Precision	Recall	Precision	Recall
Max. Sound Energy	0.85	0.75	0.77	0.90
Max. Spectral Band Energy	0.89	0.76	0.81	0.89
Low-pass Filtering	0.86	0.78	0.79	0.94
Chewing Band Power	**0.92**	0.61	**0.92**	0.87
Fractal Dimension	0.91	**0.87**	0.86	**0.98**

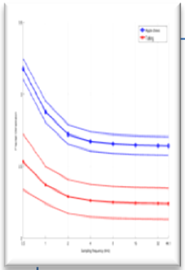
Chewing Bouts	Similar accuracy with Chewing Band Power
	Significant improvement at recall
Snacks	Relatively lower accuracy than Chewing Band Power, still better than other algorithms
	Almost 100% recall

Conclusions



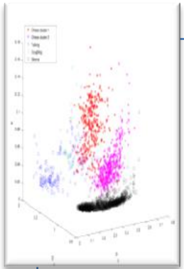
Fractal Nature

- Chewing sounds are highly fractal
- Estimation of Fractal Dimension requires only few data points (6 banks of dilation/erosion)



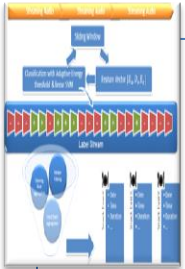
Discrimination

- Chewing sounds exhibit higher Fractal Dimension than other sounds captured by in-ear microphones
- This property persists even for severely down sampled recordings



Classification

- Fractal Dimension & Energy can be used to detect chewing activity
- Indication for food texture prediction



Algorithm

- High effectiveness in challenging dataset
- Improvement compared to literature approaches

Thank you