Fractal Nature of Chewing Sounds

Vasileios Papapanagiotou[1], **Christos Diou**[1], **Zhou Lingchuan**[2], **Janet van den Boer**[3], **Monica Mars**[3], **Anastasios Delopoulos**[1]

[1]Multimedia Understanding Group, Electrical and Computer Engineering Department, Aristotle University of Thessaloniki, Greece
[2]CSEM SA, Landquart, Switzerland
[3]Division of Human Nutrition, Wageningen University, Netherlands

Introduction

Monitoring dietary behaviour in every-day conditions can be a powerful tool in **early detection of risk** for diseases such as Obesity or Eating Disorders (e.g. Anorexia Nervosa), as in the case of the SPLENDID project (`http://splendid-program.eu/`).

A modality for monitoring dietary behaviour is the **acoustic signal captured by a microphone** conveniently placed near the outer inner ear canal. Amft et al. [1] introduced this setup, and suggested an algorithm based on estimation of statistical features of audio, and a classification scheme. Päßler et al. [2] has systematically tested seven algorithms on a large dataset.

In this work, we explore the **Fractal Dimension of chewing sound signals** and signal derivatives, and investigate its potential as a discrimination attribute against other sounds captured by such a microphone, such as talking, coughing, and ambient noise.

Fractal Dimension of Chewing Sounds

Fractal Dimension by Mandelbrot, where $A_B(\epsilon)$ is the area resulting from dilating the graph by ϵB, is defined as

$$D = 2 - \lim_{\epsilon \to 0} \frac{\log A_B(\epsilon)}{\log \epsilon}$$

For discrete signals, the area A_B can be approximated using M banks of the dilated and eroded versions of the audio segment [3]

$$A_B(\epsilon) \approx \sum_{n=0}^{N-1} \left[x_k^d(n) - x_k^e(n) \right], \epsilon = \epsilon_0 k, k = 0, 1, 2, 3, \ldots, M$$

We estimate the Fractal Dimension as

$$D = \frac{1}{M} \sum_{\epsilon=1}^{M} \frac{\log(A_B((k+1)\epsilon_0)) - \log(A_B(k\epsilon_0))}{\log(k+1) - \log(k)}$$

To examine the fractal nature of chewing sounds, we plot in Fig. 1 the data points $\log A_B(\epsilon)$ against the scaling parameter k on a set of extracted chewing and non-chewing segments from a larger dataset. The observed **linearity is a strong indication of their inherent high fractal nature**, and is **significantly higher than other sounds** such as talking, coughing, noise, etc. A maximum level of 6 banks ($M = 5$) is required to accurately estimate the Fractal Dimension of such chewing audio segments.

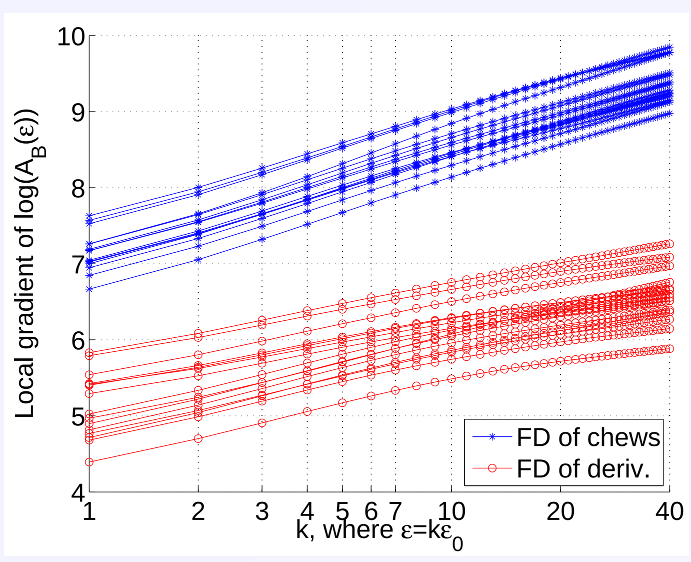

Figure 1: Local gradient for $\log A_B(\epsilon)$ versus $k = 1, 2, \ldots, 40$ (on log-scale), for 20 apple chew segments (blue) and their derivatives (red).

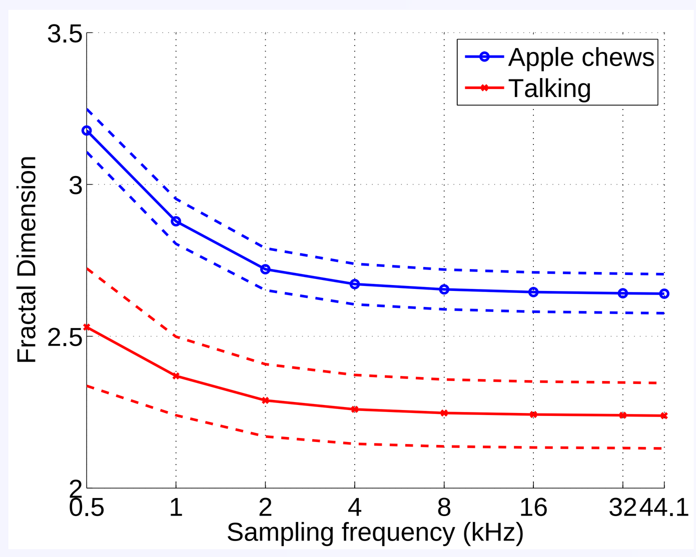

Figure 2: Mean ($\pm$ standard deviation) of Fractal Dimension of apple chews and talking segments, across various sampling frequencies (in log-scale), showing (a) the linear separability of the two classes, and (b) that down-sampling up to 2 kHz does not significantly alter the actual value of Fractal Dimension.

Down-sampling the audio signal (originally sampled at 44.1 kHz) as low as at 2 kHz does not alter the fractal nature of the chewing sounds, and can thus be used to discriminate chewing from non-chewing sounds (Fig. 2).

Forming a classifier

A feature vector of the form $[D_x, D_s, E_x]$ is estimated for each chewing and non-chewing segment of the extracted audio segments. The segment energy E_x can linearly separate segments of silence (or very low noise) from the rest of the chewing/non-chewing segments, since the Fractal Dimension of silence is not accurately estimated.

The Fractal Dimensions D_x and D_s of the audio segment and its derivative respectively can be used to discriminate chewing sounds from non-chewing ones, once the low energy segments have been eliminated (Fig. 3). It is also possible to detect clusters of various properties, such as crispy and non-crispy.

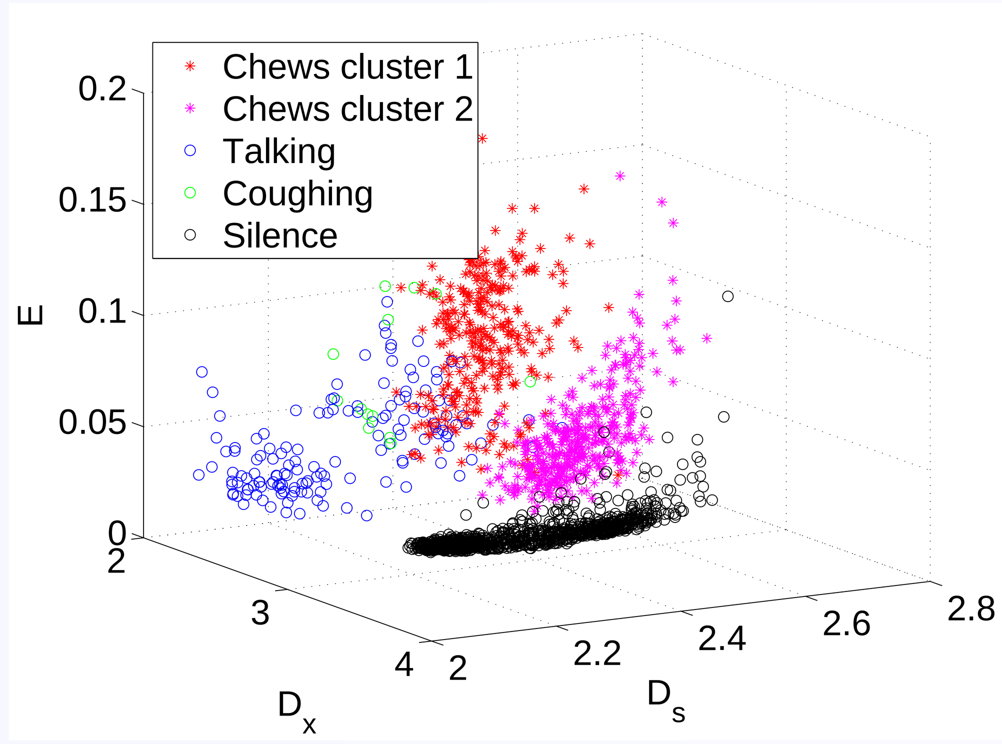

Figure 3: Feature vectors for the entire dataset, at 2 kHz. Chews cluster 1 includes banana and potato chips, and chews cluster 2 includes the remaining 4 food types.

Experiments

A dataset of 10 subjects was recorded in Wageningen University. Each subject was wearing a **prototype sensor** consisting of an FG-23329 **microphone housed in an ear bud**, designed by CSEM. Each recording lasts approximately 30 minutes, and includes a **variety of tasks** such as eating various types of foods, talking, coughing, eating in background noise/talk, etc, creating a **realistic and challenging dataset**.

Table 1: The extracted audio segments of chewing and non-chewing segments.

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

Table 2: Confusion matrix for the classification experiment with linear kernel and three classes: chew, talk/cough, and silence. Energy threshold is 0.0202, and the separating line in the $D_x \times D_s$ plane is defined by $y = -2.62x + 8.73$. Classification accuracy is 95.4%.

Class	Chew	T/C	Sil.
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

To evaluate the algorithm in real-life applications, we have combined it with an aggregation method based on adaptive energy estimation, to detect chewing bouts. We have also applied known literature algorithms using the same aggregation method for comparison.

Table 3: Precision and recall for chew bouts and snacks for the entire dataset.

	Chew bout		Snack	
Algorithm	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

Conclusions

We have performed a systematic analysis of the fractal nature of chewing sounds, which indicates that chewing sounds are **highly fractal**, compared to other sounds captured by such microphones. Thus, the Fractal Dimension can be used to **discriminate chewing from non-chewing sounds**, even in severely down-sampling the audio segments. Promising evidence was also found in discriminating between different food types (e.g. crispy/non-crispy).

Based on this evidence, we have proposed a detection algorithm and employed it to detect chewing bouts. The algorithm's computational burden is significantly low (recursive computation of dilation and erosion, very few levels of banks required to estimate the Fractal Dimension, very low bandwidth requirement of 2 kHz).

The algorithm was applied on a **challenging dataset with realistic conditions** (ambient noise, talking, coughing, etc). We have observed **significant improvement**, especially over recall, against other known counterparts of literature.

[1] Oliver Amft, Martin Kusserow, and G Troster. Bite weight prediction from acoustic recognition of chewing. *Biomedical Engineering, IEEE Transactions on*, 56(6):1663-1672, 2009.
[2] Sebastian Päßler and Wolf-Joachim Fischer. Evaluation of algorithms for chew event detection. In *Proceedings of the 7th International Conference on Body Area Networks*, pages 20-26. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2012.
[3] Petros Maragos and Alexandros Potamianos. Fractal dimensions of speech sounds: Computation and application to automatic speech recognition. *The Journal of the Acoustical Society of America*, 105(3):1925-1932, 1999.