VILNIUS UNIVERSITY
FACULTY OF MATHEMATICS AND INFORMATICS
MODELLING AND DATA ANALYSIS MASTER'S STUDY PROGRAMME

Master's thesis

Fractal modeling of speech signals

Fraktalinis šnekos signalo modeliavimas

Ieva Kizlaitienė

Supervisor: Assoc. Prof. Dr. Gintautas Tamulevičius

Vilnius, 2021

Contents

Abstract

Fractal dimension is an important characteristic of fractals that describes the degree of fractals' irregularity over multiple scales. In this paper we develop five new methods for fractal dimension calculation and analyze their theoretical and practical applications. Theoretical research is performed on extreme cases, harmonic function and Weierstrass sine function. Practical experiment uses fractal dimension-based features for speech emotion classification task. All results are compared with Katz, Castiglioni and Higuchi algorithms. The study suggests that in most cases proposed methods' performance is better and more reliable and that the new methods can be used in future researches.

Santrauka

Fraktalinė dimensija yra svarbi fraktalų savybė, apibūdinanti jų netaisyklingumą ir chaotiškumą. Šiame darbe pristatomi penki nauji metodai fraktalinės dimensijos apskaičiavimui ir analizuojamos jų teorinės bei praktinės savybės. Teorinis tyrimas remiasi ribinių atvejų, harmoninės funkcijos bei Weierstrasso sinuso funkcijos fraktalinių dimensijų skaičiavimu. Praktinis eksperimentas naudoja statistinius fraktalinių dimensijų rodiklius kalbos emocijų atpažinimui. Gauti naujų metodų rezultatai yra palyginami su Katz, Castiglioni ir Higuchi algoritmais. Daugeliu atvejų pasiūlyti metodai veikė geriau ir buvo labiau patikimi, todėl gali būti naudojami ir tolimesniuose tyrimuose.

Keywords: Fractal dimension, Speech emotion; Non-linear; Katz's method, Castiglioni's method, Higuchi's method; Neural network; Feature evaluation.

1 Introduction

1.1 Fractal Geometry and Fractal Dimension

Mathematician Benoit B. Mandelbrot was the first to introduce the concept of fractal and the fractal dimension. Mandelbrot recognized the need of new geometry as Euclidean geometry was not capable of completely describing irregular nature objects and phenomena [1]. The best explanation of standard geometry shortcomings was provided by Mandelbrot himself: "Why is geometry often described as "cold" and "dry"? One reason lies in its inability to describe the shape of a cloud, a mountain, a coastline, or a tree. Clouds are not spheres, mountains are not cones, coastlines are not circles, and bark is not smooth, nor does lightning travel in a straight line" [2]. Indeed, it is effortless to recognize that figures of classical geometry represent real-life objects rather poorly. The main contribution of Mandelbrot was his research into the findings of the earlier mathematicians and the development of a practical application of their theory. Mandelbrot introduced the term "Fractal" to describe a class of functions first discovered by Cantor (Cantor's dust), Koch (the Koch curve) and Peano [3]. Term "Fractal" was coined from the Latin word "fractus", which means "broken" and "irregular". Fractal definition relies on Hausdorff dimension d_H, which was defined in 1919 as follows:

$$d_H = \frac{logN(r)}{log(l/r)}.$$

Here bounded set $S \in \mathbb{R}^n$ is considered, $N(r)$ is a minimum number of balls of radius r required to cover the set S and l is the length of initiator. Then fractal is determined as a set with non-integer Hausdorff dimension (simply, fractal dimension). However, such definition remains rather abstract and does not help to grasp main properties of fractals. According to [4], the hard-and-fast definition of fractal does not exist, only a list of its properties. Therefore, when we refer to a set F as fractal, we typically have in mind:

1. F has fine structure (detail on arbitrary small scales);
2. F is too irregular to be described in traditional geometrical language;
3. Often F has some form of self-similarity;
4. Usually, fractal dimension is larger than its topological dimension;
5. In most cases, F is defined in a very simple way, perhaps recursively.

Here self-similarity is defined as follows: *A set $S < \mathbb{R}^n$ is called self-similar, if it is the union of disjoint subsets S_1, S_2, ... , S_n, that can be obtained from S with rotation, translation and scaling operation (-s).*

In other words, when objects with topological dimension equal to 1 are considered, fractal can be interpreted as a **geometrical curve that consists of identical shape repeating on an ever decreasing scale** [1]. Fractal dimension D is used to define the dimension of

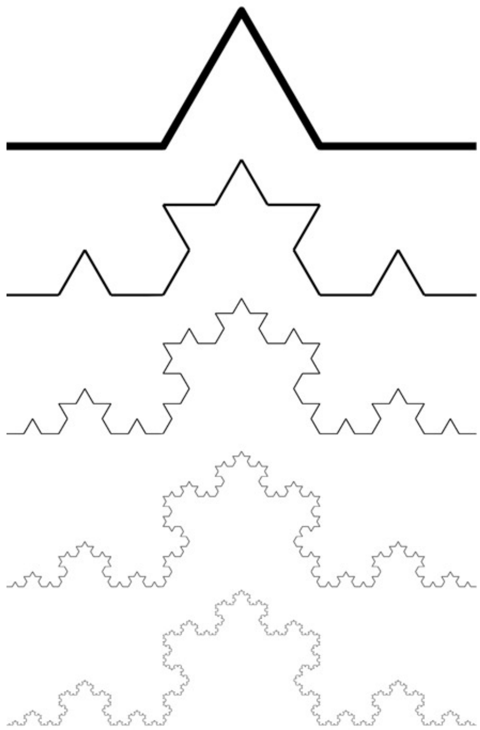


Figure 1. Fractal construction (Koch curve) [4].

some highly irregular sets and intuitively can be understood as a degree of irregularity over multiple scales [5].

Different articles ([1], [4], [6]) might provide slightly different definitions of fractal dimension (for example, $D \in \mathbb{R}$ or $D \in \mathbb{R}/\mathbb{Z}$), thus it is important to note in advance, what we consider as "Fractal dimension" in our article. The main property, on which we will rely, is that **fractal dimension characterizes how densely a fractal fills the space**. In other terms: $\mathbf{TP \leq D \leq TP{+}1}$, here TP marks topological dimension of the analyzed object. When 0 dimension object is considered (for example Cantor dust), fractal dimension cannot exceed 1, because $D \geq 1$ indicates, that our object is not simple points, but rather a curve. Similarly, when object with $TP = 1$ is considered, for example waveforms, FD should not exceed 2, as true waveforms can never be convoluted sufficiently enough to fill a plane.

Table 1. Fractal types and associated ranges of fractal dimension [7].

Fractal Dimension and Fractal types	
$0 < D < 1$	Fractal dust
$1 < D < 2$	Fractal signals and curves
$2 < D < 3$	Fractal images and surfaces
$3 < D < 4$	Fractal volumes
$4 < D < 5$	Fractal time

Another property of fractal dimensions is that they cannot be compared when measured in arbitrary units (except for ideal self-similar shapes). For instance, if the length of a planar curve is measured in millimeters, then the fractal dimension will usually be much larger compared to measurements made in kilometers.

Most articles, which are related to fractal dimension theory, emphasize theoretical properties of fractal dimension. For example, Hausdorff dimension, which is considered as the most accurate model for fractal dimension, has such properties [8]:

1. Monotonicity: $if\ E \subseteq F,\ then\ dim_H(E) \leq dim_H(F)$;

2. Finite stability: $dim_H(E \cup F) = \max\ \{\ dim_H(E),\ dim_H(F)\ \}$;

3. Countable stability : $if\ \{F_i\}_{i \in I}\ is\ a\ countable\ collection\ of\ sets,\ then\ dim_H(\bigcup_{i \in I} F_i) = sup\{dim_H(F_i) : i \in I\}$;

4. Countable sets: $if\ F\ is\ a\ countable\ set,\ then\ dim_H(F) = 0$.

However, these properties will not be analyzed in this study, because later we will focus on time series (speech signals), while theoretical approach is more based on sets. In order to apply such characteristics, we would have to 1) define subsets of speech signal (for example, express it in terms of sine functions), 2) face the problem of countability - our data will be finite and models will also take finite number of steps, while fourth property requires uncountability.

1.2 Speech Signals

Speech is the most natural way of human communication. Although for most people speech production might seem as a very simple and natural task, it is in fact a very complex process that involves a coordinated participation of several physiological structures. Speech starts with a thought in the brain, then muscular movements are activated in order to produce the sound. While speaking, the speaker constantly controls and monitors the speech production process. The vocal tract is a passageway made up of muscles and other tissues and enables production of different sounds. It also modifies the temporal and spectral distribution of power in the sound waves, which are initiated in glottis. Therefore, each utterance has its intonation, loudness, duration, and each word consists of a sequence of phonemes (units of a sound that distinguish one word from another). Speech understanding is an inverse process: final output from the speaker - signal - is passed to listener's inner ear, where frequency analysis is performed. Then follows neural transduction process and spectral signal is converted into activity signal on auditory nerve. Finally, neural activity is mapped to a language and listener is able to understand the message [9]. Here speech signal, which "travels" from speaker, is understood as an acoustic sound, and might be described by air pressure change in time (as function $f(t),\ t \geq 0$). Usually in speech signal analysis $f(t)$ is substituted by $f(t_0),\ f(t_1),\ \ldots\ ,\ f(t_N)$, as only finite number of observations can be performed. Time $\hat{t}$, at which function $f(\hat{t})$ is measured, is called discretization time, and $\Delta t = t_i - t_{i-1}$ is called discretization step [10].

Signals are classified as deterministic and stochastic. Deterministic signal's future values can be defined for indefinitely large t. For example, signal $F(t) = A \cdot sin(2\pi ft)$ is defined for all t. Here $A \geq 0$ is amplitude, which shows the signal size, and f is frequency, measured in Hertz (Hz). 1 Hz is equal to one oscillation per second, or, in other terms, a cycle of wave with frequency 1 Hz lasts for 1 second. $F(t)$ is also called a harmonic wave and is a stationary signal, as its statistical characteristics do not vary in time. Stochastic signals are random signals, meaning that their future values can not be accurately computed. Usually stochastic signals are non-stationary, as their properties might change for different t. Speech signals are non-stationary, as each sentence, word, syllable and even letter have different features.

Signals are called continuous, when time t changes continuously, and discrete, when signal is measured only at a discretization time. In this article discrete time sinusoidal signal will be analyzed, therefore, proper discretization step should be used. Our solution will be based on Nyquist–Shannon sampling theorem [10]:

If signal $f(t)$ does not contain any frequencies higher than B hertz, then $f(t)$ can be completely determined by measuring its values at a series of points spaced $\frac{1}{2B}$ seconds apart.

It is also worthy to note that sine functions are very important in speech signal analysis, because each signal can be decomposed as sums of sine waves [9].

Speech signal is treated as one-dimensional and is said to depend on one argument — time t. Assumptions, that speech is one-dimensional planar wave, have lead to comprehensive understanding on how speech sounds are produced and what their properties are. It also contributed to extensive knowledge about speech analysis, modification and modeling. However, such approximation has its drawbacks, as it does not reflect full reality. Air in the vocal tract system is not static, but moves from lungs out the mouth, carrying the sound field along with it, thus containing also a non-acoustic component [11]. While speaking, in the vocal tract turbulent excitations occur. Turbulence is a non-linear phenomena with strong interaction between the acoustic sound field and the airflow occurring at obstacles and constrictions. Then, highly irregular fluctuations of pressure and particle velocity appear. Turbulence is also the source of creaky and breathy voice, whisper, fricative sounds, thus it plays important role in creating comprehensive speech models. Modern theories, that attempt to explain turbulence, predict the existence of eddies at multiple scales [12]. According to the energy cascade theory, energy produced by eddies is transferred hierarchically to the small-size eddies which actually dissipate this energy due to viscosity. This multiscale structure of turbulence can in some cases be quantified by fractals.

1.3 Fractal Dimensions in Speech Modeling

Speech recognition is an important and well known subject that has been widely studied for many years. While in the beginning speech recognition stayed more in theoretic fields,

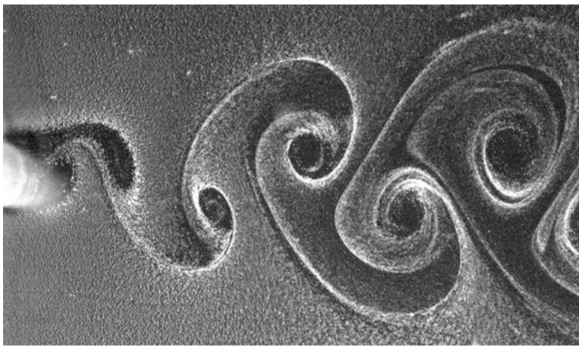


Figure 2. Sound of the wind [13].

currently we face a variety of real-world applications. Technologies, based on speech recognition, started to change our lives and became a commonness for many of us. The most popular application is virtual assistants (*Siri*, *Google*, *Alexa*, etc). Now we are able to search for information, make purchases, play music, send emails and messages using only our voice. There exist even more use cases: voice biometry can be used for security improvement, transcribing interviews, pod-casts, work meetings, voice recognition is even used for catching criminals, etc. For example, *Apple* released *HomePod* speaker: "When you say "Hey Siri", advanced signal processing, together with echo and noise cancellation, allows *HomePod* to hear you without the need to raise your voice — even if you're across the room with loud music playing" [14]. The *HomePod* is also capable of recognizing and distinguishing individual speaker from other people, which is also an important task in speech recognition field.

One of the possible ways of deeper speech analysis is investigating its non-linearity. Theoretical background, provided in subsection 1.2, motivates the use of fractals in this paper. In order to measure the "fractalism" of speech signals, or, in other words, in order to quantify the degree of signal's graph fragmentation, fractal dimension is used. However, the relationship between fractal geometry (or fractal dimension) and turbulence is currently very little understood, thus in this Master thesis we conceptually equate the amount of turbulence in a speech sound with its fractal dimension.

2 Fractal Dimension Algorithms

2.1 Introduction

The concept of fractal dimension originates from Hausdorff dimension d_H, which we already defined in Chapter 1.1. A detailed study regarding the analytic properties of Hausdorff dimension was mainly developed by Besicovich and his pupils during XX century [8]. Hausdorff dimension is admitted to be the most accurate method for fractal dimension. However, its definition is quite abstract, thus it is difficult or sometimes even impossible to calculate in practical applications. Therefore, various studies and researches were made, proposing a bunch of different methods and algorithms for fractal dimension calculation. Most of the methods have their theoretical and/or practical limitations and often give different results.

2.2 Box Counting method

Quite often fractal dimension is understood as classical **box counting dimension**. Box counting method, which can also be called the grid or reticular cell counting, is so far the most popular algorithm due to its effective calculation and successful results. In addition, box counting method satisfies monotonicity and finite stability, which were defined in Chapter 1.1. Prior to box counting method applications in fractal dimension calculation, it was used for evaluating area of irregular cartographic features. Below we provide steps of basic implementation:

1. Cover the analyzed object/feature with one rectangle (box);
2. Divide box into four quadrants and count the number of occupied cells;
3. Divide each subsequent quadrant into four sub-quadrants and continue doing so until the minimum quadrant size is equal to the resolution of the data. Each time keep track of number of cells occupied.
4. Mark n as number of filled boxes (cells), b as box size, and plot $log\ n$ against $log\ b$. Then the slope with opposite sign is equal to fractal dimension [15].

Currently Box counting method is mostly used in digital media research field, such as analyzing digital shorelines and contours (Fig. 3), modeling river networks, inspecting photographs of vegetation, biological images (for example it was applied in revealing Retinal vasculature, investigating naevus and melanomas), etc. [15–17]. Due to the fact that applications are concentrated in 2D image analysis, in this article we decided to exclude Box counting method from our scope, as speech signals are 1D objects. Therefore, we will analyze algorithms, which intend to estimate planar curves fractal dimension.

Figure 3. Box counting method [6].

2.3 Katz algorithm

In 1987 M. J. Katz published his article proposing a new way to estimate fractal dimension of waveforms [18]. His idea was that fractal dimension can be computed as follows:

$$D_{Katz} = \frac{log(\frac{L}{a})}{log(\frac{d}{a})}, \tag{1}$$

here L is the total length of the curve, d is the diameter (planar extent) of the curve and a is the average distance between two successive points. Step a is used as a normalization factor, which discards the impact of the signal dynamic range. In our models a will be the discretization step which we defined in Chapter 1.2. Noting, that $L = n \cdot a$, where n is the number of steps in the curve, we can rewrite (1):

$$D_{Katz} = \frac{log(\frac{L}{a})}{log(\frac{d}{a})} = \frac{log(n)}{log(d) - log(\frac{L}{n})} = \frac{log(n)}{log(d) - (log(L) - log(n))} = \frac{log(n)}{log(n) - log(\frac{d}{L})} \tag{2}$$

Length of the curve L can be interpreted as accumulated change of speech signal values and is calculated using simple Euclidean distances formula:

$$L = \sum_{i=1}^{n} l_{i,i+1} = \sum_{i=1}^{n} \sqrt{1 + (y_{i+1} - y_i)^2} \tag{3}$$

Planar extent d is equal to the maximum distance between the first point and any other point of the curve:

$$d = max\{l_{1,i}\},\ 2 \leq i \leq n + 1. \tag{4}$$

From equation 2 we see that when length increases, i.e. when a curve becomes more dense and/or its amplitude grows, then $log(\frac{d}{L})$ increases, thus the denominator decreases and fractal dimension is larger. And vice versa, when L drops (curve becomes more straight, less dense), D also decreases.

Katz fractal dimension answered the purpose in many different articles, sometimes as the main algorithm, evaluating the fractal dimension [19], sometimes giving additional results for comparisons and conclusions [20–23].

2.4 Castiglioni algorithm

Despite successful applications, Katz algorithm also received a lot of criticism. In 2010, Paolo Castiglioni, Italian researcher in biomedical technology field, wrote a letter to the editor of journal "Computers in Biology and Medicine", named "What is wrong in Katz's method?" [24]. In this letter, he provided comments regarding the article "A note on fractal dimensions of biomedical waveforms" [20]. P. Castiglioni's letter was publicly acknowledged due to his reasonable critics and proposal for improvement. The main drawback, which was emphasized, is low values of Katz dimension. Even when applied to Brownian motions with true fractal dimension $1.0 - 1.5$, Katz algorithm gives almost a constant value very close to 0.

Castiglioni suggested slightly modifying Katz algorithm and currently his modification is known as Castiglioni algorithm. He argued that it is misleading to take into account horizontal axes, even after normalization, as time t and measured quantity y are intrinsically different. Therefore, the distance between pairs (t_k, y_k) and (t_i, y_i) is actually meaningless and might distort the final result. Consequently, Castiglioni proposed to apply Mandelbrot's approach directly to mono-dimensional space y_k and remove time dimension from extention d and length L calculations:

$$d = max\{y_k\} - min\{y_k\} \tag{5}$$

$$L = \sum_{k=1}^{n} |y_{k+1} - y_k| \tag{6}$$

Final expression of Castiglioni fractal dimension is the same as (2):

$$D_{Cast} = \frac{log(n)}{log(n) - log(\frac{d}{L})}. \tag{7}$$

However, Castiglioni method also has some weaknesses. It is easy to notice that the extent d is vulnerable, when curve has outlier(-s) with big or small y_k value(-s). In such cases, $log(\frac{d}{L})$ might increase significantly and thus D_{Cast} might also grow substantially. A more detailed analysis of this drawback will be presented in the next chapter.

On the other hand, Castiglioni's remark is considered as meaningful and important, thus in most applications his method goes together with Katz method, providing additional insights and results [22, 23, 25, 26].

2.5 Higuchi algorithm

In 1988, Japanese scientist T. Higuchi published the article "Approach to an irregular time series on the basis of the fractal theory", where he introduced a new efficient technique to measure fractal dimension of the set of points $(t, f(t))$ [27]. Higuchi method turned out to give stable characteristic time scale and trustworthy fractal dimension. His method is presented below:

First of all, let's mark the observation $f(t)$ values as $X(1), X(2), \dots, X(N)$ at respective discretization times $t_1, t_2, \dots, t_N$, where N stands for number of observations and step size

Δ is fixed for all t. Then new time series X_k^m are constructed:

$$X_k^m = \{X(m), X(m+k), X(m+2k), \dots, X(m + \lfloor m,k \rfloor \cdot k)\}. \tag{8}$$

Here $k \in [1, 2, \dots, k_{max}]$ indicates a discretization period, $m \in [1, 2, \dots, k]$ is a starting index value and $\lfloor m,k \rfloor = \lfloor \frac{N-m}{k} \rfloor$ ($\lfloor a \rfloor$ means integer part of a). In other words, m can be understood as the initial time and k can be understood as the interval time. For example, if $N = 100$ and $k_{max} = 2$, we will get three different sets X_1^1, X_2^1 and X_2^2:

1. $k = 1, m = 1, \lfloor m,k \rfloor = \lfloor \frac{100-1}{1} \rfloor = 99$

   $X_1^1 = \{X(1), X(1+1), X(1+2\cdot 1), \dots, X(1+99\cdot 1)\} = \{X(1), X(2), X(3), \dots, X(100)\}.$

2. $k = 2, m = 1, \lfloor m,k \rfloor = \lfloor \frac{100-1}{2} \rfloor = 49$

   $X_2^1 = \{X(1), X(1+2), X(1+2\cdot 2), \dots, X(1+49\cdot 2)\} = \{X(1), X(3), X(5), \dots, X(99)\}.$

3. $k = 2, m = 2, \lfloor m,k \rfloor = \lfloor \frac{100-2}{2} \rfloor = 49$

   $X_2^2 = \{X(2), X(2+2), X(2+2\cdot 2), \dots, X(2+49\cdot 2)\} = \{X(2), X(4), X(6), \dots, X(100)\}.$

Note that for each k we get k different time series, whose union covers all observations. Then, the frequency properties of each resampled signal are estimated:

$$L_m(k) = \frac{N-1}{\lfloor m,k \rfloor \cdot k^2} \cdot \sum_{i=1}^{\lfloor m,k \rfloor} |X(m+ik) - X(m+(i-1)k| \tag{9}$$

Here term $\frac{N-1}{\lfloor m,k \rfloor \cdot k}$ represents the normalization factor for the curve. Further the length of the curve for the time interval k is defined:

$$\overline{L(k)} = \sum_{m=1}^{k} L_m(k) \tag{10}$$

If $\overline{L(k)}$ is proportional to k^{-D}, then the curve, describing the shape of the epoch, has dimension D. Therefore, if $\overline{L(k)}$ is plotted against $k = 1, 2, \dots k_{max}$ on a double logarithmic scale, the points should fall with a slope equal to $-D$. In short, if $\overline{L(k)}\ \alpha\ k^{-D}$, then curve's fractal dimension is equal to D.

The main uniqueness of Higuchi method is evaluation of properties regularity over different time scales. $\overline{L(k)}$ stability for different k shows that curve is not irregular and gives lower value of D. On the contrary, if analyzed time series are highly irregular, then resampling will result in quite different versions of the original signal and D will also increase.

It is important to note that the choice of the free parameter k_{max} has a significant impact on final results. However, Higuchi himself did not extensively analyze the problem of optimal k_{max} selection. Moreover, he used much greater k_{max} (for example 2^{11}) compared to what other authors later used. Several articles were published trying to evaluate the best k, however, optimal solution for all tasks was not found [28–30]. Currently, in most articles, maximum k does not exceed 20 and is chosen as fixed value by each author. In this Master thesis, similar approach will be used.

Higuchi algorithm proved to be a very good numerical approach for rapid assessment of signal non-linearity (currently his paper is cited almost 900 times). There exists a variety of different applications: recognition of speech signals [23], researches of animals behavior [22], many different use cases in medicine field, such as analysis of biomedical waveforms [20, 29], EGG, neuron and MEG signals [21, 28], Alzheimer disease patterns [25], etc. [26, 30, 31].

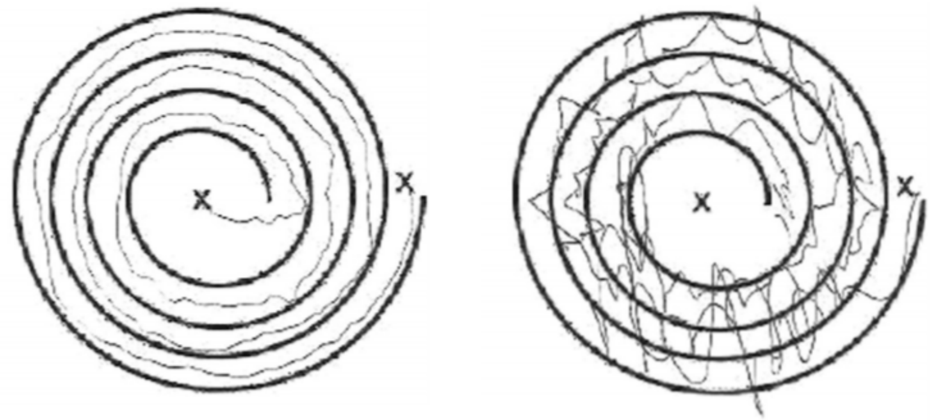


Figure 4. Parkinson disease progress evaluation from Archimedes spiral drawings [26].

2.6 Other

In fact, there exist much more different methods for fractal dimension estimation. Due to the limited scope of this thesis, we decided to perform detailed analysis and comparisons only with Katz, Castiglioni and Higuchi algorithms. The decision was also based on the fact that these methods are the most common in speech signal analysis. Therefore, in this chapter we will mention only a few other popular methods together with short descriptions.

2.6.1 Divider method

The Divider method was first used for calculating the length of cartographic lines. Later mathematicians recognized that this method could be easily applied in fractal dimension estimation. Method's idea is simple — a chosen divider is "walked" along the object contour and number of needed steps is recorded. Divider's width is systematically increased and, after repeating the "walking" process, the relation between the step size and line length over a range of resolutions is determined [15]. Applications of this algorithm are mostly related to various digitalized paths, contours analysis, for example investigating animal movement paths [32] or doing researches on geomorphic phenomena [33].

2.6.2 Hurst exponent

Hurst exponent H is used as a measure of long-term memory of time series. It can be understood as the "index of long-range dependence" and it helps to estimate data series randomness. It assesses how the variability of the data changes with the length of the time being considered. The originally proposed technique for evaluating H is based on a rescaled-range analysis of time series. The ratio of range and standard deviation is measured for all possible length segments of the analyzed data [15, 23]. Applications of this method are concentrated in time series analysis field. For example, exploration of Foreign Exchange Markets [34], forecasting of animals' population [35], or even analysis of speech signals [23, 36].

2.6.3 Petrosian's algorithm

Petrosian's method provides a fast computation of the fractal dimension of the time series by translating the data into a binary sequence. Final formula is similar to Katz and Castiglioni algorithms: $D_{Petrosian} = \frac{log(n)}{log(n)+log(\frac{n}{n+0.4N_{\Delta}})}$. Here n stands for number of measured signal values and N_{Δ} is the number of sign changes in the binary sequence. This binary sequence is formed by assigning 1 for each difference between consecutive points when it exceeds a standard deviation magnitude. Otherwise 0 is assigned. [37]

2.6.4 Sevcik's algorithm

Sevcik's method idea is to subject waveform to a double linear transformation that maps time series to a unit square. The abscissas and ordinates are normalized and then fractal dimension is obtained using the following formula: $D_{Sevcik} = 1 + \frac{ln(L)+ln(2)}{ln(2N')}$. Here L is the length of the time series in the unit square and N' stands for number of measured signal values minus 1 [38].

3 Methodological part

3.1 Katz, Castiglioni and Higuchi methods' theoretical properties

The aim of this sub-chapter is to show that currently developed algorithms are not perfect and that there is room for improvements and new creative ideas. Evaluation was made using three different approaches: we checked extreme cases, analyzed algorithms' behavior on simple harmonic function and investigated methods' results of computing fractal dimension of Weierstrass sine function, which has a known theoretical FD.

3.1.1 Extreme cases

Three different extreme cases (A, B, C) were analyzed:

(A) $f(t) = 0 \ \ \forall\, t \in [0, 1]$;

(B) $f(t) = t \ \ \forall\, t \in [0, 1]$;

(C) $f(t_i) = \begin{cases} 1, & \text{if i is odd} \\ -1, & \text{if i is even,} \end{cases}$, here $i > 0$ marks order of points t.

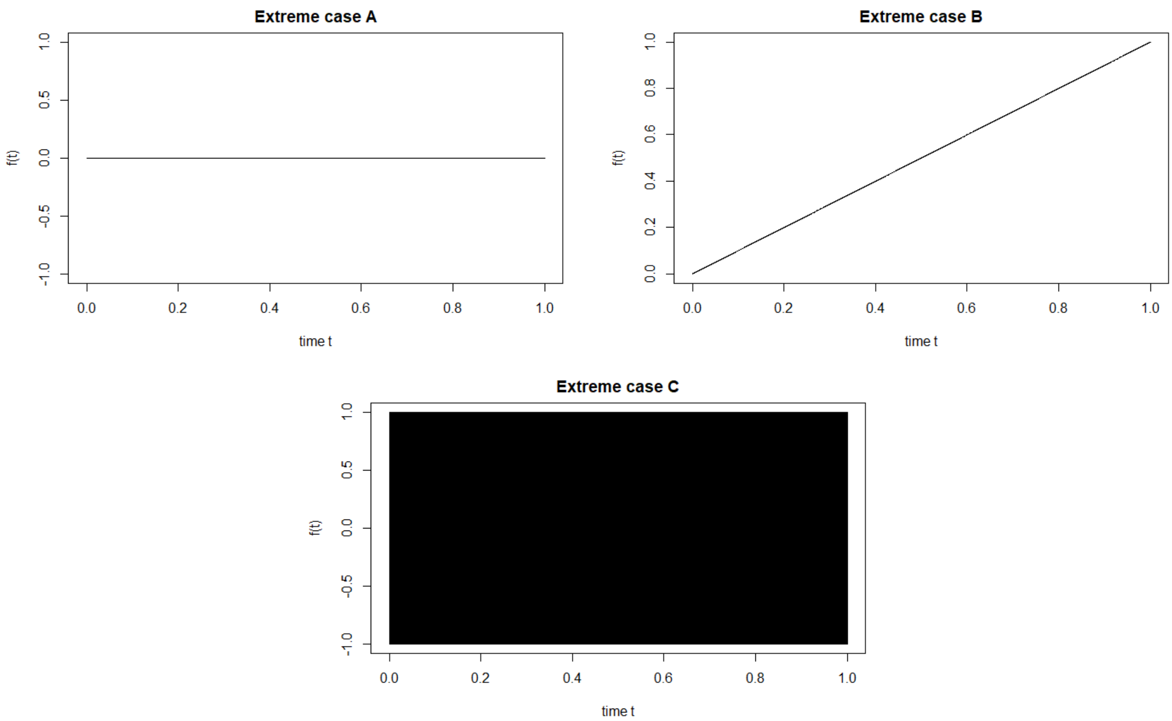


Figure 5. Extreme cases.

Results are presented in the Table 2 below:

Table 2. Estimated fractal dimension of extreme cases.

FD	Katz	Castiglioni	Higuchi	Expected value
A	1	NaN	1	1
B	1	1	1	$1+\varepsilon$
C	1.1	Inf	2.5	2

It is easy to notice that Katz gives too low values, as in case C fractal dimension is estimated to be only 1.1. On the contrary, Castiglioni method gives too high value for case C (dimension diverges to infinity). Castiglioni is also not capable of evaluating case A, as there is no change in amplitude. Higuchi overestimates C value, but is the closest to expected values. This analysis shows, that all algorithms are not always reliable and perfect.

3.1.2 Function $A \cdot sin(2\pi ft)$

In order to properly evaluate the method, we decided to analyze harmonic function $f(t) = A \cdot sin(2\pi ft)$ with parameters $A = 1$, $t \in [0,1]$, $f = 50 \cdot i$, $i \in [0,20]$. Number of steps n was fixed at 5000 and as $f_{max} = 1000$, Nyquist–Shannon sampling theorem conditions were satisfied.

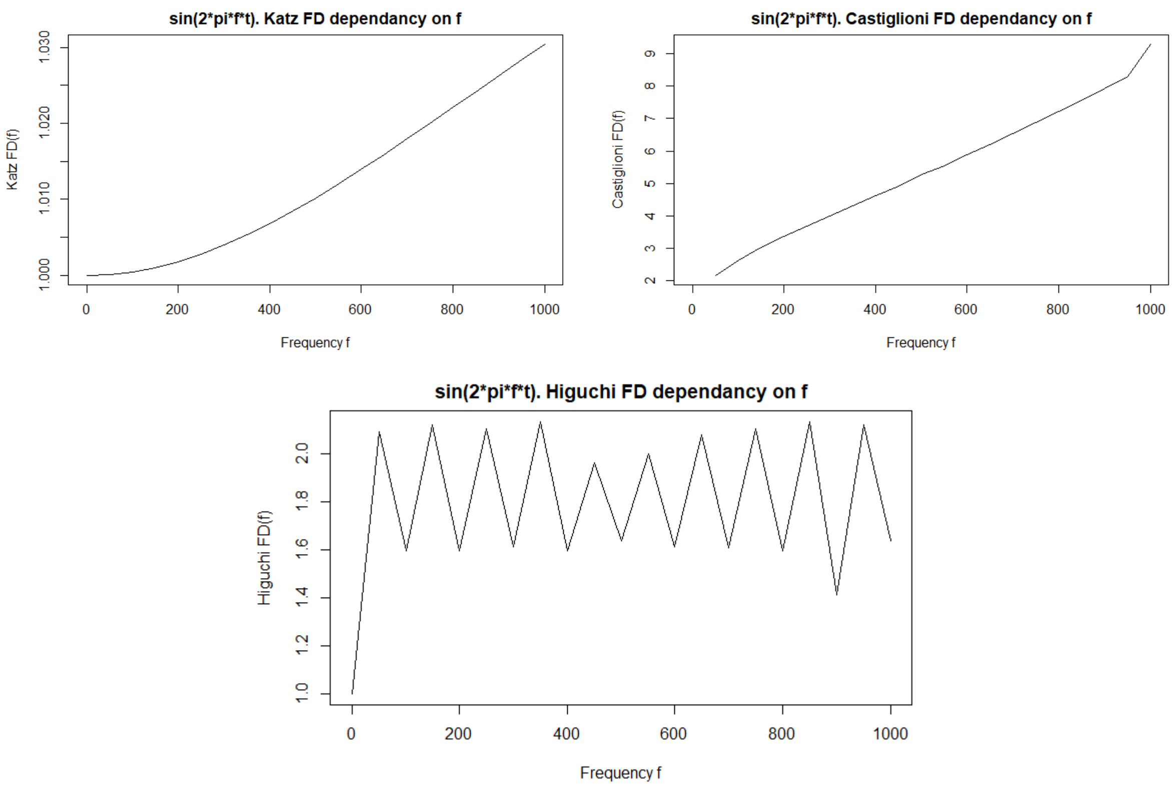


Figure 6. $sin(2\pi ft)$ fractal dimensions.

Results turned out to depend strongly on the analyzed method:

- Katz, as previously, gives too low values. It slowly increases as f grows, but when frequency is 1000 it still gives only ~ 1.03, whereas the expected output logically should be ~ 1.5 or higher.

- Castiglioni, similarly to Katz, grows as f increases. However, values are too high: already with $f = 50$ it estimates fractal dimension ~ 2, and consistently increases finally reaching 9.

- Higuchi behaves exclusively: it does not maintain linear dependency on frequency. The main reason behind fluctuations is our step size relation to periodicity: in some cases more irregularity over time is recorded , while in other cases step size might "catch" similar function values and thus record lower irregularity.

3.1.3 Weierstrass sine function

Harmonic function analysis shows some trends of methods properties, but does not provide any strictly proven arguments about the reliability of these methods, as we do not know the real fractal dimension of a sine function. Therefore, we decided to investigate a function, which has a theoretical fractal dimension - Weierstrass sine function:

$$f(x) = \sum_{j=1}^{\infty} \lambda^{-\alpha j} sin(2\pi \lambda^j x),\ 0 < \alpha < 1,\ \lambda > 1.$$

It is the most famous example of everywhere continuous but nowhere differentiable function, which was discovered by Karl Weierstrass. He himself managed to prove non-differentiality only for some λ and α, but later it was proved for all λ and α values [39]. The graph of this function (together with some other Weierstrass functions) has been studied as an important example of fractal curves and its fractal dimension was proved to be $2 - \alpha$, when $\lambda > 1$.

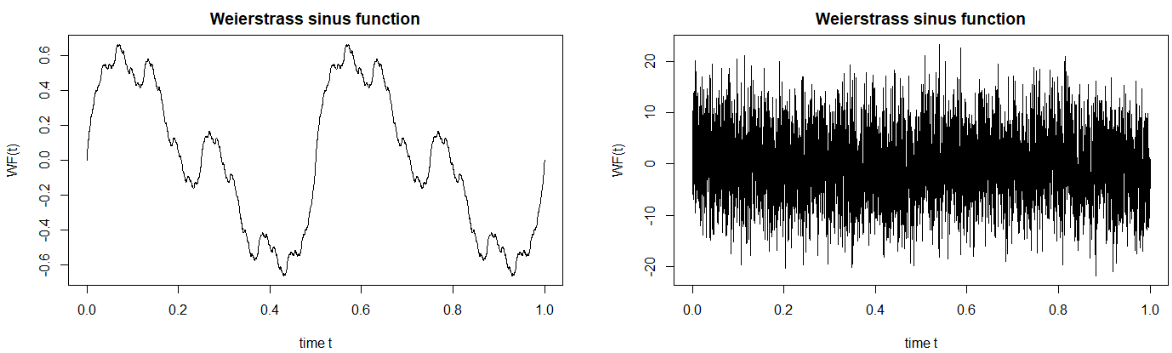


Figure 7. Weierstrass function, $\lambda = 2$, $\alpha = 1$ on the left, $\alpha = 2$ on the right.

In this article we examined function

$$f(x) = \sum_{j=1}^{\infty} 2^{-\alpha j} sin(2^{j+1} \pi x),\ \alpha \in [0,\ 0.1,\ 0.2, ..., 1.9,\ 2].$$

Final results were as expected: Katz values are too low, Castiglioni - too high, and Higuchi seems to almost match theoretical dimension.

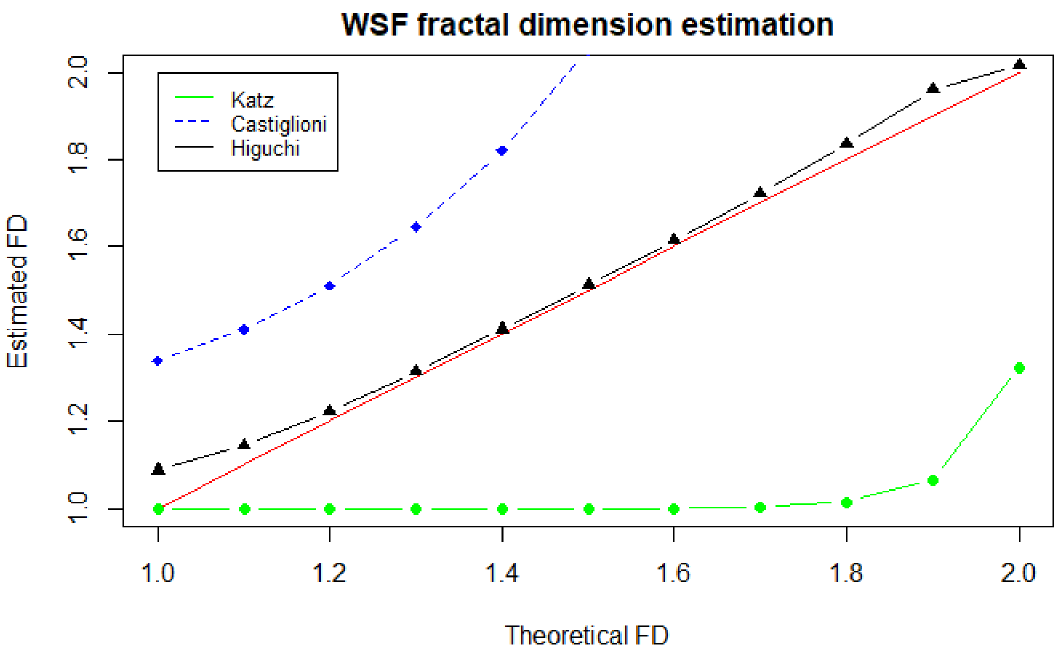


Figure 8. WSF fractal dimension.

3.2 Proposed methods

The aim of this thesis is to create new improvement or new algorithm for fractal dimension calculation. Our proposals are described in the following subsections.

3.2.1 Amplitude Fractal Dimension

Amplitude method intends to evaluate the ratio between current amplitudinal changes and all possible amplitudinal changes. After identifying the maximum and the minimum signal values, we make an assumption, that the biggest fractal dimension value would be obtained if in each step signal changes its value from maximum to minimum or from minimum to maximum. It would guarantee fully filled space (as in extreme case C). Then this hypothetical curve's length is calculated and the ratio between this length and the real length is computed:

$$L = \sum_{i=1}^{n} |y_{i+1} - y_i|$$

$$d = max\{y_i\} - min\{y_i\},\ i \in [1,\ n],\ l = d \cdot (n-1)$$

$$D_{ampl} = 1 + \frac{L}{l}, \tag{11}$$

here n marks number of signal values.

3.2.2 Distance Fractal Dimension

Distance method is similar to Amplitude method, but it has two adjustments:

- Horizontal axes are taken into account;

- Discretization time is divided into subsets and d is calculated for each subset values separately.

$$L = \sum_{i=1}^{n} \sqrt{\Delta^2 + (y_{i+1} - y_i)^2}$$

$$k := \lfloor \frac{n}{10} \rfloor$$

$$X_i = \{X(1 + (i-1) \cdot 10),\ X(2 + (i-1) \cdot 10),\ \ldots\ X(10 + (i-1) \cdot 10)\},\ i \in [1, (k-1)]$$

$$X_k = X / \{X_i\}$$

$$d_j = max\{X_j\} - min\{X_j\},\ l = 10 * \sum_{j=1}^{k-1} \sqrt{\Delta^2 + d_j^2}$$

$$D_{diff} = 1 + \frac{L}{l}, \tag{12}$$

This algorithm is more comparable to Katz, as both, x and y, axes participate in distance calculations. A split by subsets helps to avoid outliers and peaks in the graph — if one signal value is significantly higher or lower than others, hypothetical curve's length drastically increases. This problem is relevant in Katz and Castiglioni algorithm as well. We could have applied the same approach to our "amplitude" method, however, different tactics were kept purposely, as we want to test various approaches.

As Amplitude and Distance methods are quite similar, we decided to apply logarithms to one of these methods. As Distance method already has a "boost" from supplementary interval splits, we adjusted Amplitude method (11) and transformed it to:

$$D_{ampl} = 1 + \frac{log_{10}(L)}{log_{10}(l)}, \tag{13}$$

3.2.3 Sign Fractal Dimension

Sign method idea came from the goal to evaluate the meandering and the waviness of the curve. We decided to try a straightforward algorithm: compare the number of local maximum and minimum points to the maximum hypothetical number of local extremum points.

$$D_{Sign} = 1 + \frac{A}{n-1},\ \text{here } A \text{ is nmbr of local max/min points.} \tag{14}$$

3.2.4 LR Intersection Fractal Dimension

Name LR Intersection stands for "Linear Regression Intersection" and aims to evaluate curve's irregularity by computing number of intersection points with fitting linear regression. The basis of this idea is the expectation that more erratic curve will have more intersection points and vice versa.

Let $y = at + b$ be a fitting linear regression of a curve $f(t)$ and let $A := \sum_i 1$, where i satisfies the condition $y(t_i) = f(t_i)$. Then

$$D_{LR} = 1 + \frac{A}{n-1} \tag{15}$$

3.2.5 Poli Intersection Fractal Dimension

As Linear Regression model might be insufficient to represent all irregularities (for example if curve has preeminent peaks), we decided to also try polynomial regression and, similarly as in LR Fractal dimension, compute intersection points. A chosen polynomial degree is 18, as it gives decent approximation and is efficient enough.

Let $y = a_{18}t^{18} + a_{17}t^{17} + \ldots + a_1 t^1 + a_0$ be a fitting polynomial regression of a curve $f(t)$ and let $A := \sum_i 1$, where i satisfies the condition $y(t_i) = f(t_i)$. Then

$$D_{Poli} = 1 + \frac{A}{n-1} \tag{16}$$

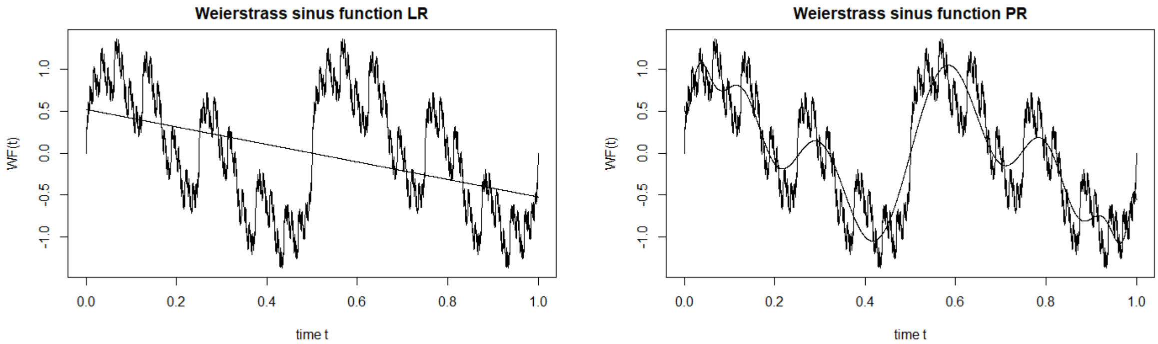


Figure 9. Weierstrass sine function. Linear and Polynomial Regressions.

Similarly to Amplitude and Distance methods, LR and Poli algorithms are also alike. Therefore, in order to have more comprehensive results, logarithms were applied to LR method 15:

$$D_{LR} = 1 + \frac{log_{10}(A)}{log_{10}(n-1)} \tag{17}$$

3.3 Behavior of new methods

The aim of the following subsections is to analyze theoretical properties of the new methods.

3.3.1 Extreme cases

Methods, decribed in 3.2 subsection, were applied on extreme cases. Results are presented in the table below:

Table 3. Estimated fractal dimension of extreme cases.

FD	Amplitude	Distance	Sign	LR	Poli	Expected value
A	NaN	2	1	1	1	1
B	1.5	1.11	1	1.18	1.27	$1 + \varepsilon$
C	2	2	2	2	2	2

The outcome is improper only in few cases:

- Amplitude method gives undefined value in Extreme case A, as there is no change in amplitude at all;
- Amplitude method value in case B is too high due to logarithms (without them, value would be 1.0002);
- Distance method gives value 2 in case A, because there is no amplitudical changes, therefore the biggest hypothetical distance (which is equal to 1) is fully covered by the line.

We cannot truly evaluate the success of Distance, Sign, LR and Poli methods in B case, as true dimension is unknown. We expect it to be close to 1 or equal to 1. But overall results are closer to expected values compared to Katz, Castiglioni and Higuchi outcomes (2 Table).

3.3.2 Function $A \cdot sin(2\pi f t)$

For further properties' analysis estimated fractal dimension of a harmonic function is investigated. The results together with the dependency on f are represented in the following graph:

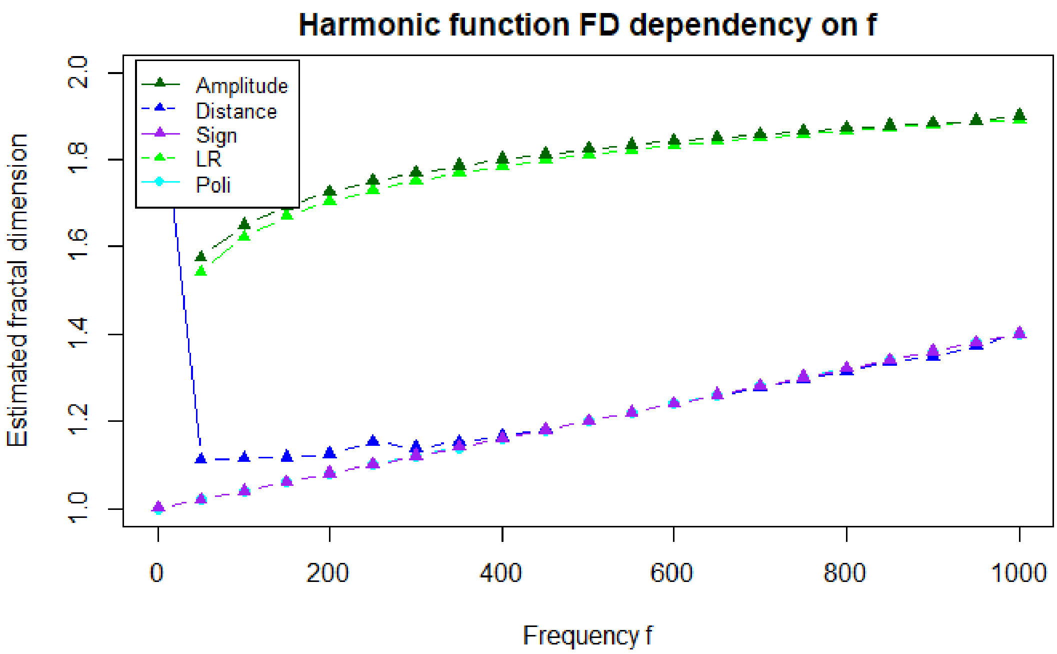


Figure 10. $sin(2\pi ft)$ fractal dimension. New methods application.

As previously, f changed from 0 to 1000. When $f = 0$, we have extreme case, similar to already analyzed case A. From the graph we can note that:

1. All methods give similar results (just Amplitude and LR algorithms are logarithmized) due to perfect periodicity of sine function.

2. All methods follow the expected pattern - they increase when f grows;

3. All methods manage to gain higher values than Katz and lower values than Castiglioni, which is a positive result. The trend is also more logical than Higuchi, and values when f is small are closer to reality (except Distance method in extreme case).

Analysis of the harmonic function shows that the methods are promising. However, we cannot prove their reliability with this approach, as true sine function fractal dimension is unknown and our evaluation is based on intuitive understanding.

3.3.3 Weierstrass sine function

The final step in theoretical investigation is applying algorithms on function, whose fractal dimension is known, - Weierstrass sine function. Results are presented in the graph below:

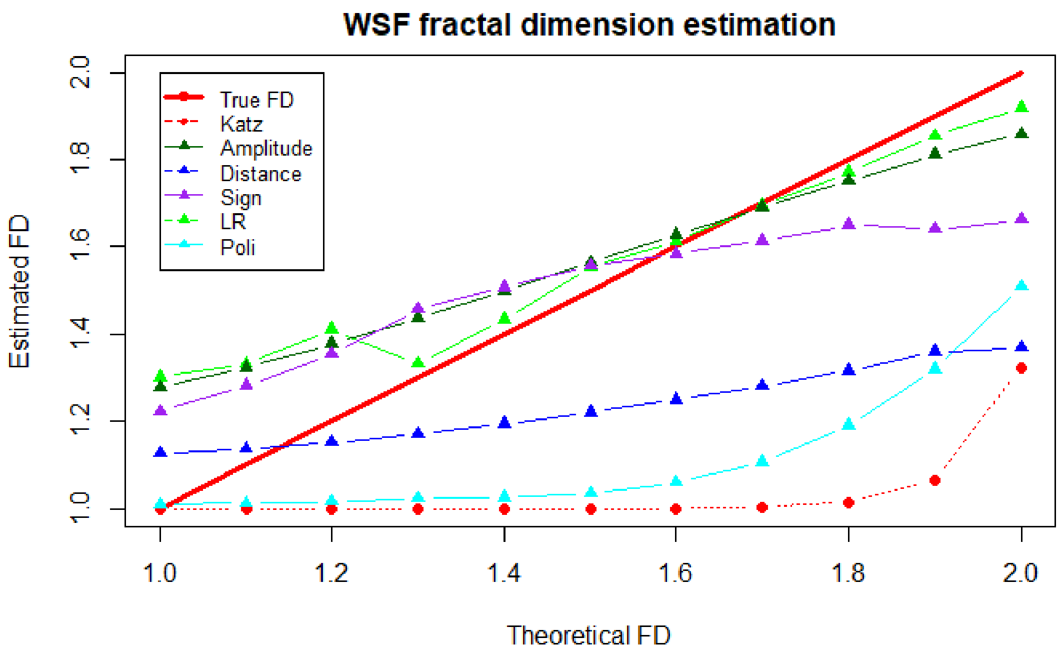


Figure 11. WSF fractal dimension. New methods application.

Katz method values were added intentionally (even though they, together with Castiglioni and Higuchi results, were already presented in 8 example), in order to have a convenient comparison with other algorithms. Therefore, after analyzing the graph, we can state:

1. Poli method behaves very similarly to Katz method, but gives slightly higher values, which proves better reliability of the method;

2. Amplitude and LR methods look rather similar and show the best results thanks to the application of logarithms;

3. Sign method, which basis is not similar no any of the methods, also shows very good results;

4. All methods follow the correct trend - increase, when true fractal dimension grows;

5. All methods seem to be more trustworthy than Castiglioni, also on average are closer to reality than Katz. However, Higuchi algorithm results remain exemplary and edifying;

6. All methods, except Poli, has higher values when α is close to 1. The reason behind this is that Weierstrass sine function is somewhat irregular even when $\alpha = 1$. It contradicts the assumption, that such function has fractal dimension 1, as we expect that simple line (for example extreme case A) also has fractal dimension of 1. Similar comment can be made when α is close to 2 - function is highly irregular, but certainly does not fully fill the space.

Overall results are promising and satisfactory, as the aim of these methods is to improve speech analysis in comparison with Katz and Castiglioni methods, which are widely used. Promising results motivate the usage of these methods in practical tasks. Applications on real-life data will be presented in the next chapter.

4 Experimental part

Theoretical reliability is a very important part of fractal dimension algorithms. Nevertheless, in this paper we intend to consider both, theoretical and practical advantages of the proposed methods. Our first step in experimental part is to examine the ability of the proposed methods to visually distinguish different sounds (e.g., vowels and consonants, some groups of consonants, or one specific sound). Second step is fractal dimension application for speaker emotion identification. Categorization task will be carried out, supplementing the full picture of new methods' utility.

4.1 Distinguishing different sounds

The aim of this small experiment is to visualize fractal dimensions of different sounds. Our data consists of 23 different sounds, each having 12 recordings (the total number of utterances — 276). For each voice recording we calculated fractal dimension using 5 different methods — Amplitude, Distance, Sign, LR and Poli. Using *Orange* platform we plotted the results (figures 12, 13, 14, 15, 16), labeling each sound and using different colors for vowels and consonants (yellow and dark blue respectively).

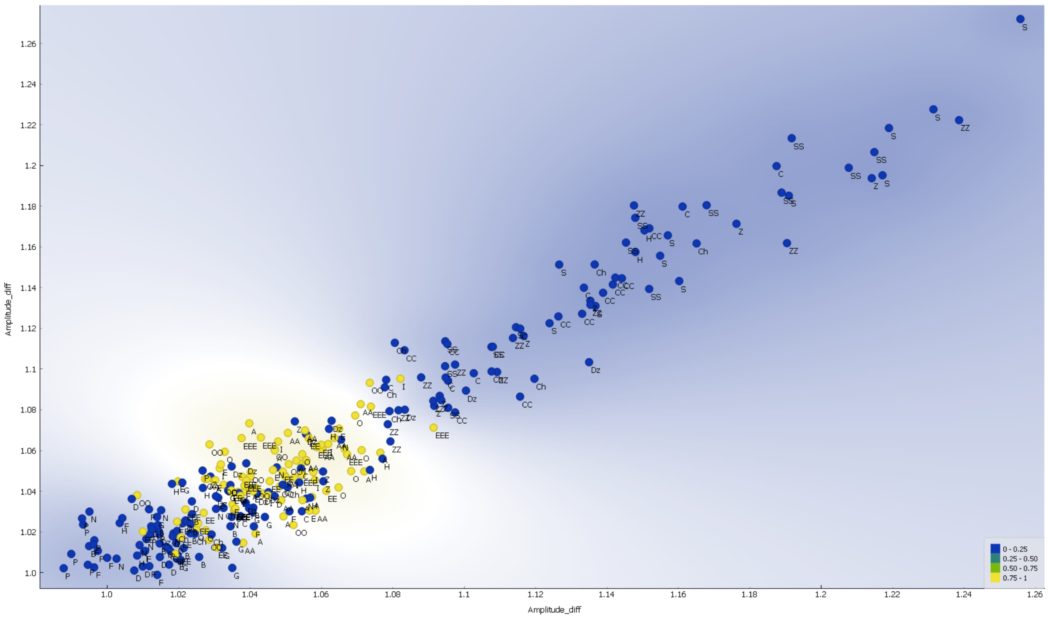


Figure 12. Fractal dimension value scatter plot: Amplitude method.

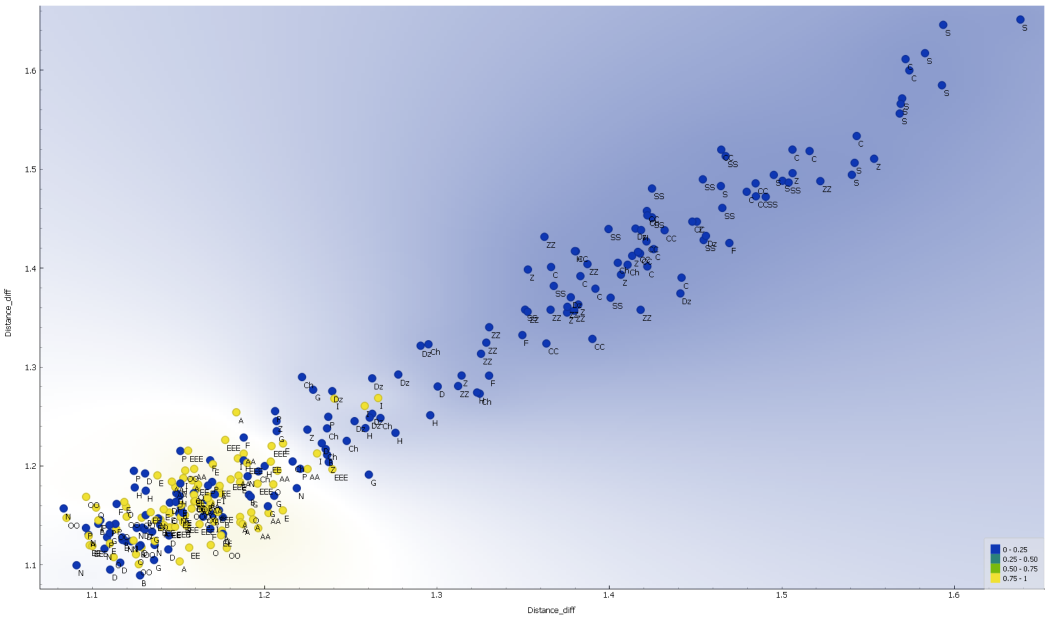


Figure 13. Fractal dimension value scatter plot: Distance method.

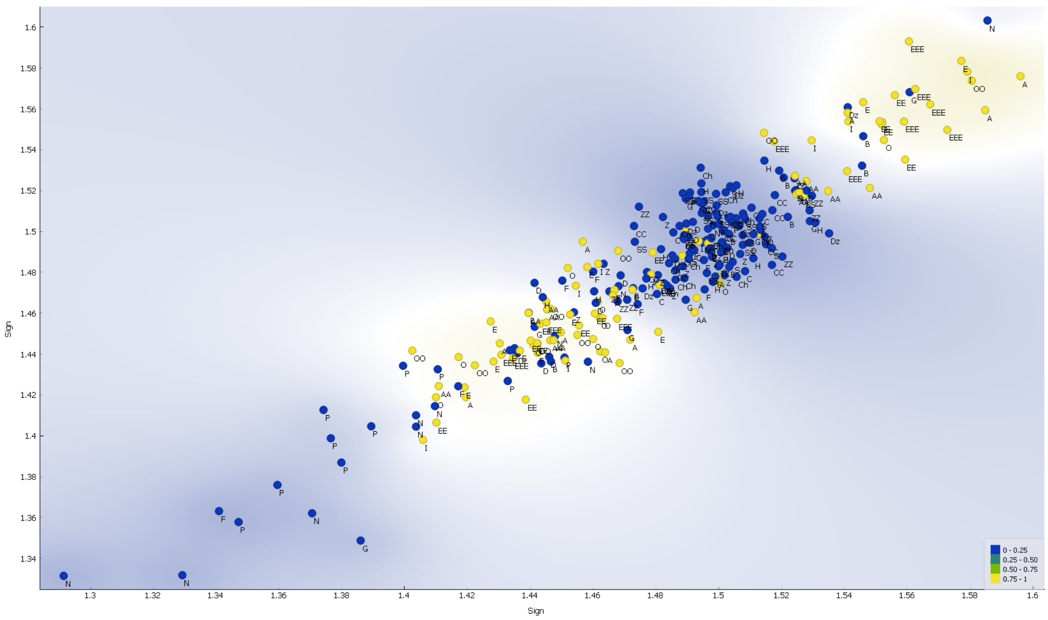


Figure 14. Fractal dimension value scatter plot: Sign method.

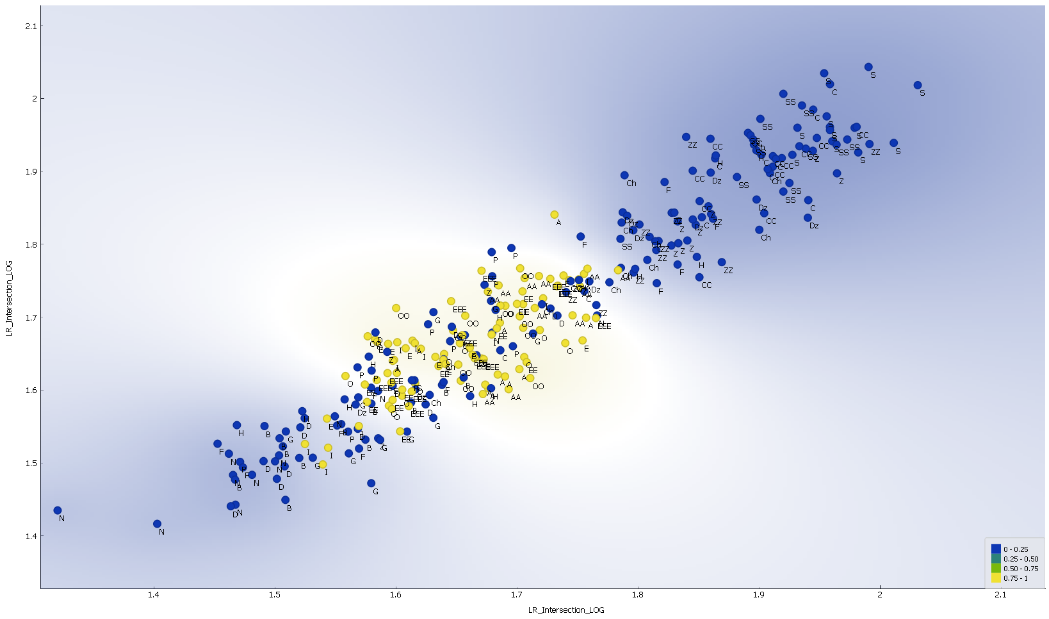


Figure 15. Fractal dimension value scatter plot: LR method.

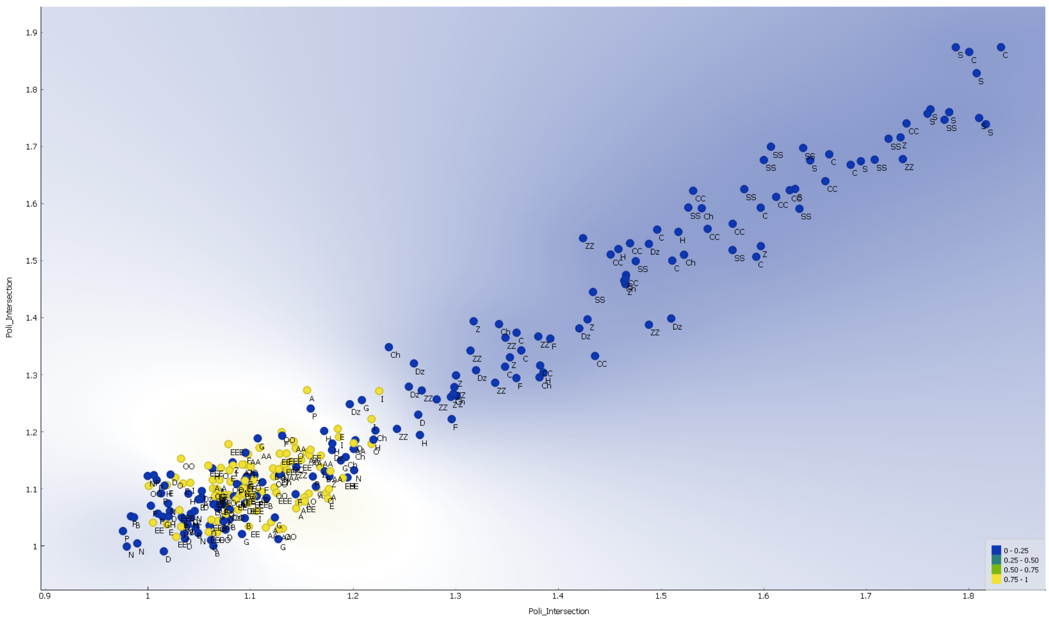


Figure 16. Fractal dimension value scatter plot: Poli method.

The graphs demonstrate that:

- Results obtained from Amplitude, Distance, LR and Poli algorithms are similar: they quite well separate vowels and consonants and are able to highlight several sounds, such as “s”, “z”, “ss” (long “s”), “zz” (long “z”), “c” and “dz”;
- The outcome from Sign method turned out to be different: vowels and consonants were not that well separated and no particular sounds group was identified.

The conclusion from this small research is that fractal dimensions have potential for discriminating different sound signals and new methods can be useful in more complex tasks. However, they should be carefully and thoroughly examined as not all of them might work properly (for now it seems that Sign method could be unreliable).

4.2 Speaker emotion identification

The aim of this experiment is to use new fractal dimension estimation algorithms in emotion recognition. The data set of sentences, pronounced in three different emotions, will be analyzed. The goals are as following:

1. To create a model for emotion classification by applying different methods;
2. To evaluate the Accuracy of different algorithms and to compare new versus already existing;
3. To explore combinations of different methods;
4. To provide additional insights, discovered during the experiment.

4.2.1 Data preparation

The analyzed data set consists of 3000 observations. Its structure:

- 3 emotions: Joy (J), Neutral state (N), Sadness (S);
- Speaker number (1-10);
- Gender (5 Female (F), 5 Male (M));
- Recording.

In total 100 different sentences were recorded, each speaker repeated these sentences 3 times (with 3 different emotions). Spoken language - Lithuanian. Recordings length is not fixed and depends on each speaker's individual pace. It is important to note that emotions are artificial, acted by each speaker by his/her own interpretation.

The emotional state impacts the entire process of speech production: respiration, muscle activity in larynx, vibration of vocal chords, dryness of the mouth, etc. These factors result in different frequency, loudness, duration of sounds and pauses and other speech emotion markers. Figure 17 demonstrates one speaker's different emotions while pronouncing a sentence "A lot of effort is needed in order to make a high quality job" (in Lithuanian: "Kokybiškam darbo atlikimui reikia įdėti daug pastangų").

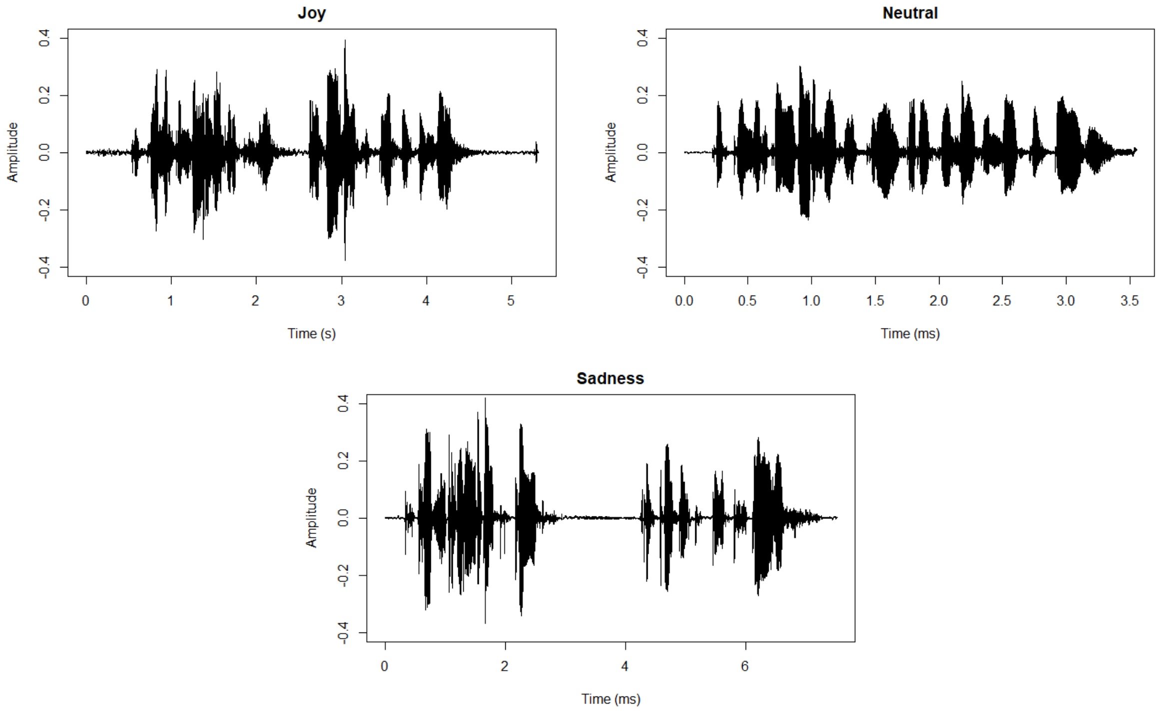


Figure 17. Same sentence. Different emotions.

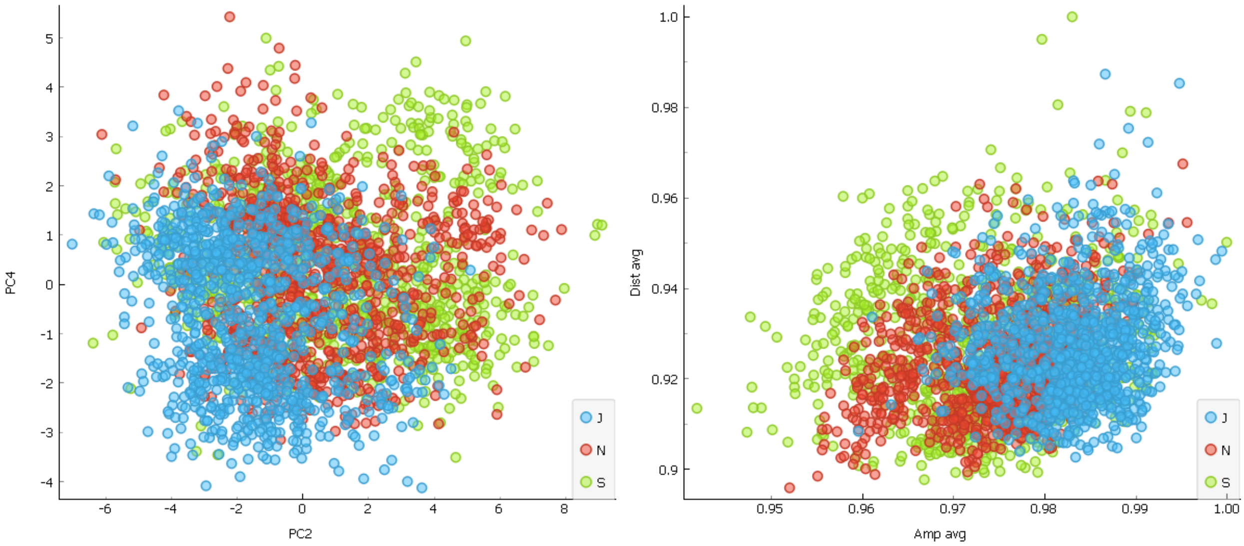
As one sentence has different words, syllables and sounds, it would be inefficient to compute one fractal dimension for the whole recording. Therefore, recordings were split into 20 ms intervals and for each interval fractal dimension was calculated using 8 different algorithms: Amplitude, Distance, Sign, LR, Poli, Katz, Castiglioni and Higuchi.

The result of this step was eight fractal dimension value sequences for each emotional utterance. Second step was establishing utterance-level features using statistical generalizations of FD sequences. The following statistical functionals were computed: arithmetical mean (marked as "avg"), geometrical mean (marked as "geom avg"), variance (marked as "variance"), maximal value (marked as "max"), minimal value (marked as "min") and mode (marked as "mode"). In total, a full set of 48 utterance-level features was extracted for each recording. Considering different value ranges and variation of the calculated statistical values, all extracted features were normalized using the maximum value of each feature.

4.2.2 Visualizing the data

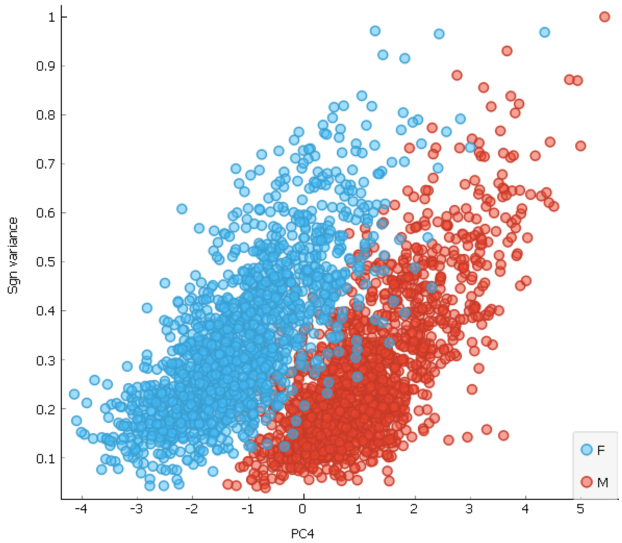
The final data set was analyzed using *Orange* platform. First of all we were interested in the general view of our models values. In order to do that, principal components were calculated from new methods statistical values (the total of $5 \cdot 6 = 30$ dimensions). 4 principal components covered 67% of variance and, unfortunately, it was not enough to represent separated sounds (Fig. 18, on the left; PC2 and PC4 were chosen as *Orange* best suggestion). In addition, the best projection (suggested by *Orange*) without Principal Components is presented (Fig. 18, on the right).

Figure 18. Observations' scatter plot



However, PCA component helped to nicely demonstrate gender split (Fig. 19 on the left, axes values were chosen as *Orange* best suggestion).

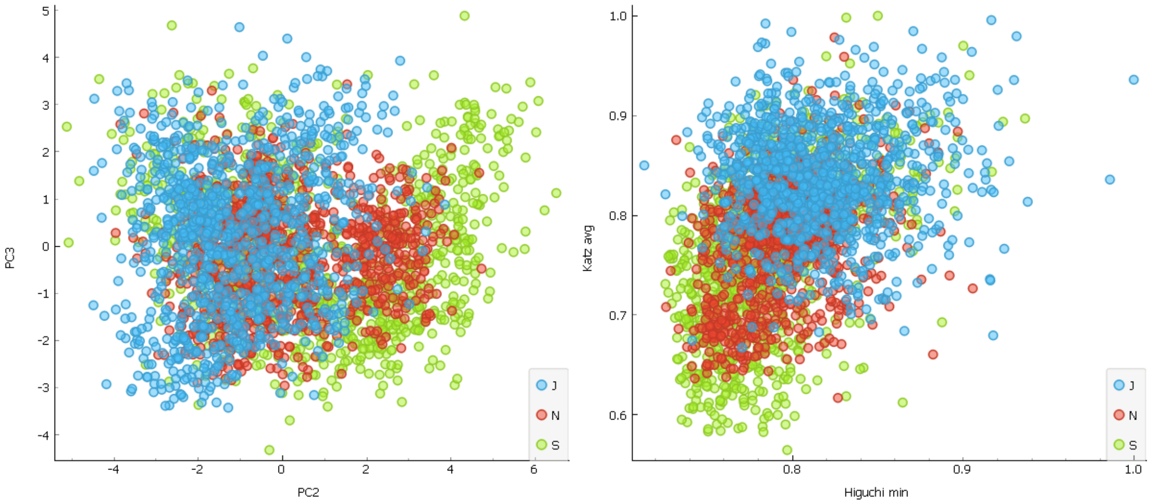
Figure 19. Observations' scatter plot



So far it seems that calculated values overlap a lot and it might be difficult to achieve prosperous results. On the other hand, at least some differences can be seen only from two features, thus, combining more features could result in better outcomes.

Before going deeper into classification models, we also estimated currently existing methods (Katz, Castiglioni, Higuchi) visual projections. Similarly as with new methods, we extracted principal components and used *Orange* suggestion for visualization (20 fig., on the left) and pictured values without principal components using just two features, suggested by *Orange* (Fig 20, on the right).

Figure 20. Observations' scatter plot



From presented graphs we can state that neither method is capable of separating emotions immediately just from the values without any classifiers. In order to compare new versus already existing methods, we have to carry out the classification task.

4.2.3 Data classification

Emotion recognition models are based on widgets available in *Orange* platform (Fig. 21 shows used scheme). In "file" widget full final data in .csv format is loaded. Then it is passed to a scatter plot (with and without extraction of principal components) for exploratory analysis purposes. The main part of the model begins with "select columns" widget, were wanted columns (methods, features) are selected. Here the target variable — emotions — is chosen. Light pink widgets — "Neural network", "Naive Bayes" and "kNN" — are classifiers: methods, which will learn how the data should be classified. We decided to use 3 different classifiers for better comparison and accuracy purposes.

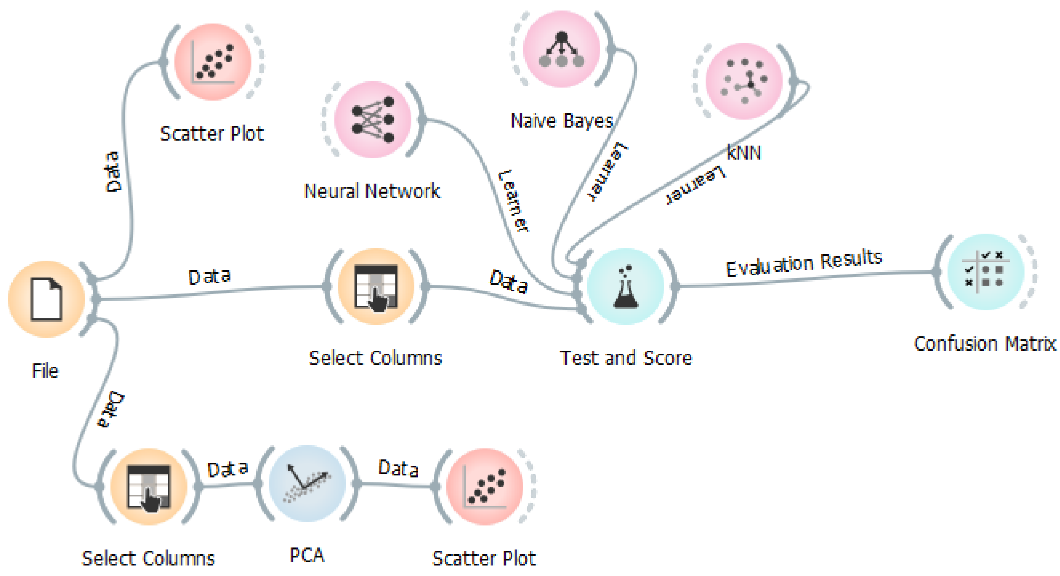


Figure 21. Experimental model in *Orange* platform.

Neural network widget in *Orange* uses a multi-layer perceptron (MLP) algorithm with back-propagation [40]. The input to Neural Network is the data, which comes from the "Test and Score" widget. The output is a trained model, or, in other words, a learning algorithm. Neural Network widget has several parameters, which we have chosen after many trials:

- Number of hidden layers (in this model - 2);
- Number of neurons in each hidden layer (in this model - 20 and 12);
- Activation function: Identity, Relu, tanh, Logistic (in this model - Identity);
- Solver: Adam, SGD, L-BGFS-B (in this model - Adam);
- Regularization parameter α (in this model - 0.0001);
- Maximal number of iterations (in this model - 10000).

The second trainer is kNN widget. The idea of the input and output is the same as in Neural network, just the working principle is different: kNN model is based on distances between training instances [40]. It searches for k closest training examples in the feature space and uses their average as prediction. Available parameters:

- Number k of neighbors (in this model - 6);
- Metric: Euclidean, Manhattan, Chebyshev, Mahalanobis (in this model - Euclidean);
- Weight: Distance or Uniform (in this model - Distance).

The third chosen trainer is Naive Bayes, which is a simple and fast probabilistic classifier based on Bayes' theorem with the assumption of feature independence [40]. The input and output is the same as in Neural Network and kNN widgets, but in Naive Bayes widget no parameters are available.

Two last widgets ("Test and Score" and "Confusion Matrix") are chosen for evaluation of emotion classification accuracy. In "Test and score" widget different sampling techniques can be applied. In this paper, a 10-fold cross-validation scheme was chosen. In addition, this widget also provides calculated main evaluation results: Area under receiver-operating curve (AUC), classification accuracy (CA), weighted harmonic mean of precision and recall (F1), the proportion of true positives among instances classified as positive (Precision) and proportion of true positives among all positive instances in the data (Recall). While comparing different methods and features, we decided to apply the criterion of the CA. The final widget, "Confusion Matrix", which shows proportions between the predicted and the actual class.

Our first step was the evaluation of different methods CA by different classifiers using all available features (Fig. 22). "All new" mark the combination of all **proposed** methods' features.

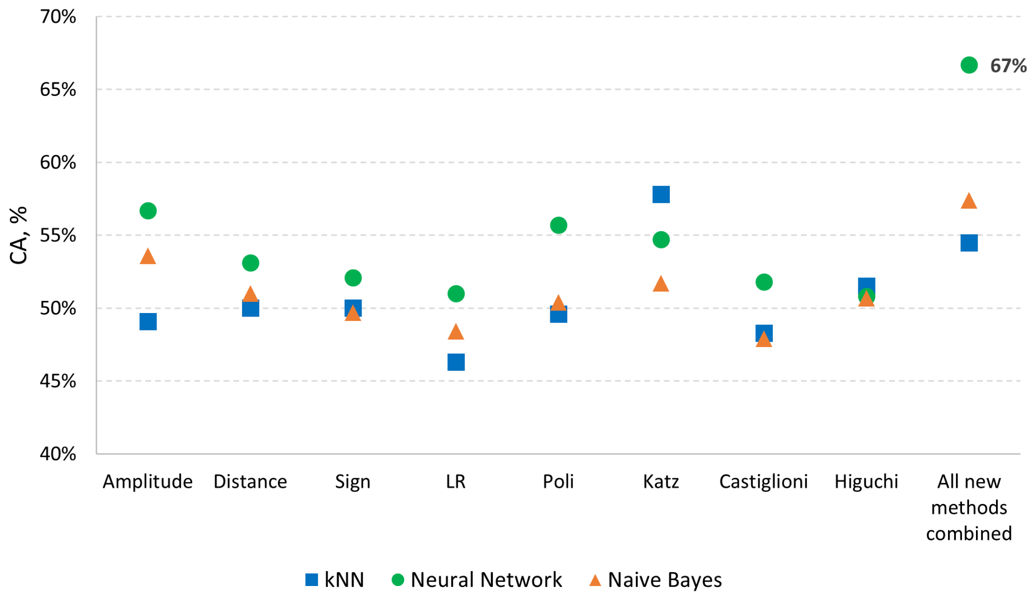


Figure 22. Emotion classification accuracy by different methods.

Achieved accuracy differs by both the method and the classifier. The best result is obtained by a combination of proposed algorithms with neural network classifier — 67%. The second best CA is reached with Katz method — 58% (kNN). Only a little worse performance comes from Amplitude and Poli methods (57% and 56% respectively). In addition, Distance and Sign methods manage to overtake Castiglioni and Higuchi methods (looking at best classifier value), while the worst method with 51% CA is LR. From this graph we can surely state that proposed methods are able to compete with currently existing methods, reaching similar or even higher results. Moreover, the combination of proposed methods outstands all other CA results, proving the value of our new ideas.

The next plot represents the accuracy, obtained by different statistical functionals, when combining different **proposed** methods (Fig. 23).

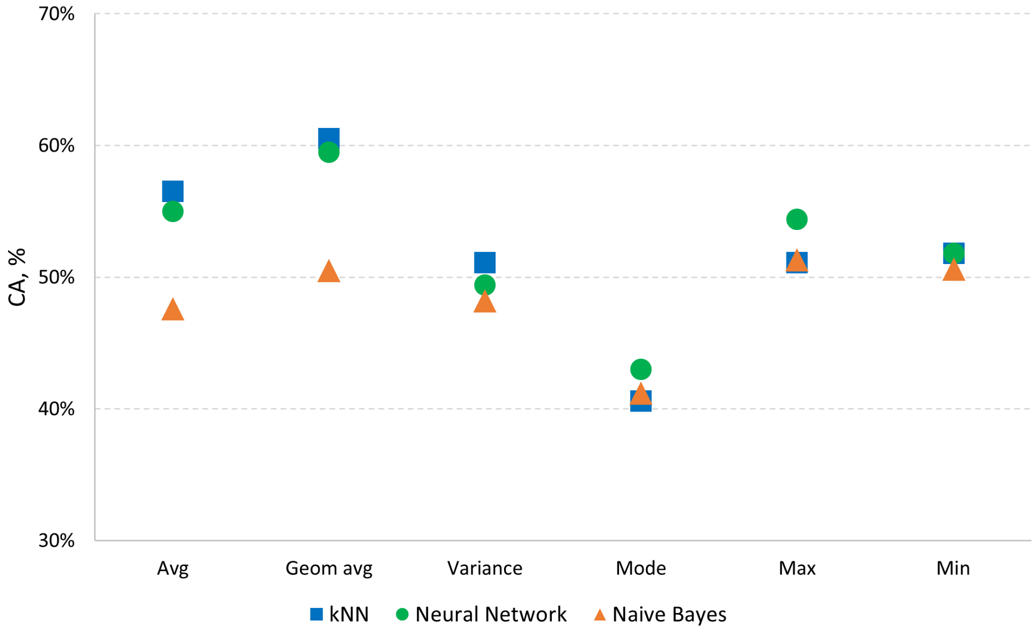


Figure 23. Emotion classification CA by different features (all new methods combined).

The purpose of this analysis is to understand whether all statistical parameters work similarly or not. From the chart it is seen that accuracy strongly depends on parameter. The most advantageous feature is geometrical average, but it alone does not reach the max achieved CA rate (61% versus 67%). The most inefficient feature is mode, but the CA rate is at least higher than random selection rate 33%. This graph motivates the need of combination of different features - some methods might perform better without one or several statistical parameters (for example without computed mode). For such feature selection we decided to use a simple trial method and, in order to apply it, we calculated each feature's accuracy, when it is used alone in classification task (Fig. 24, 25, 26). For comparison purposes we also added a dashed line, which represents each method's accuracy rate when all features are used.

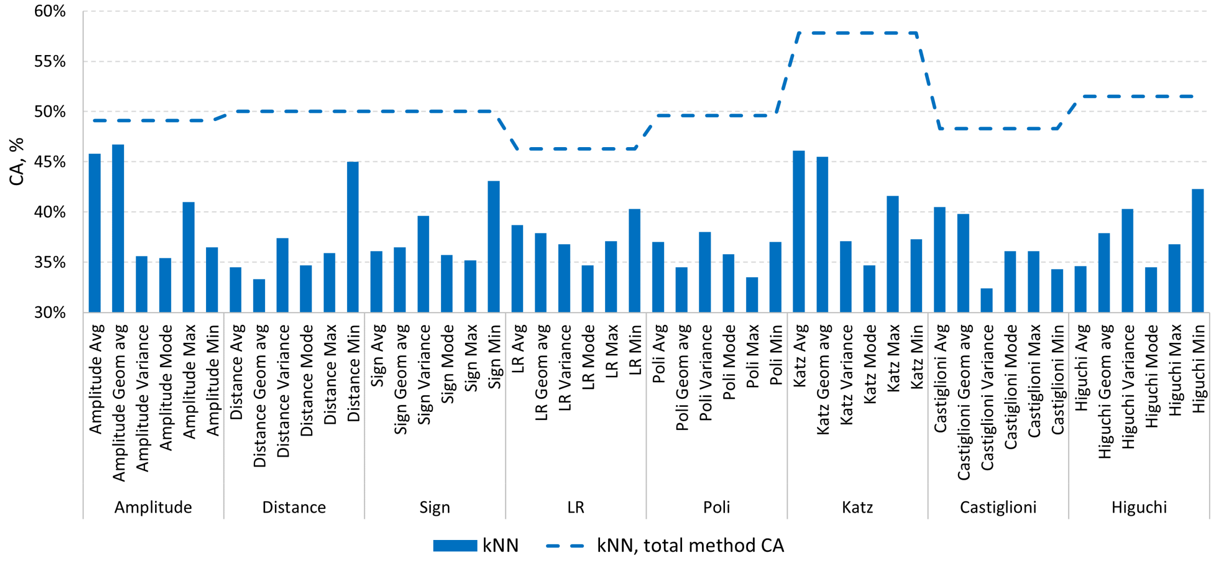


Figure 24. CA by features, kNN method.

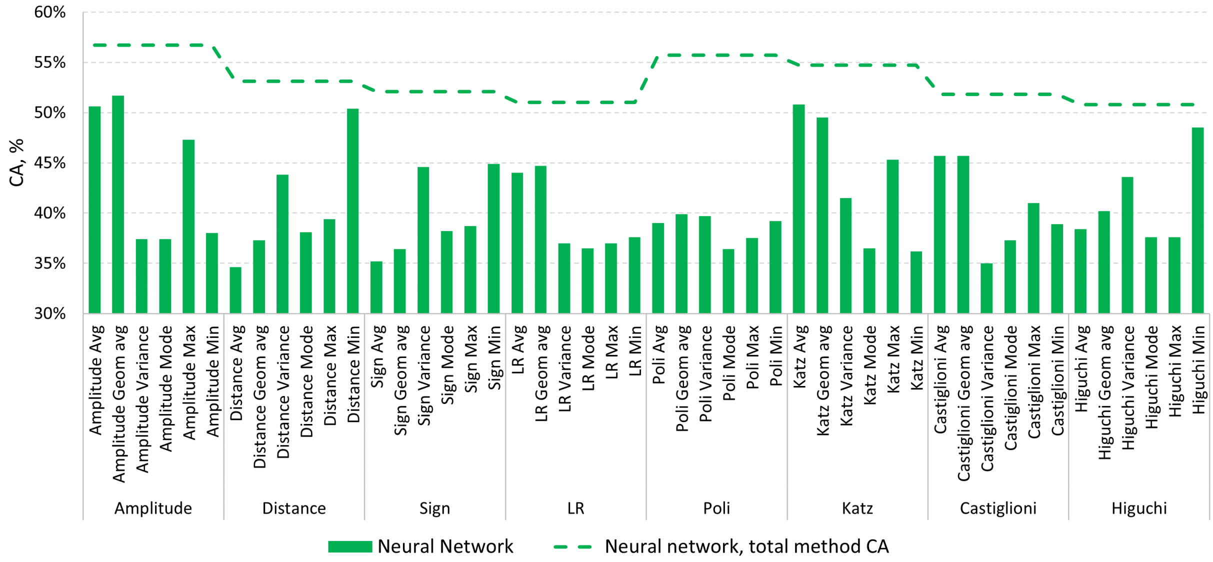


Figure 25. CA by features, Neural network method.

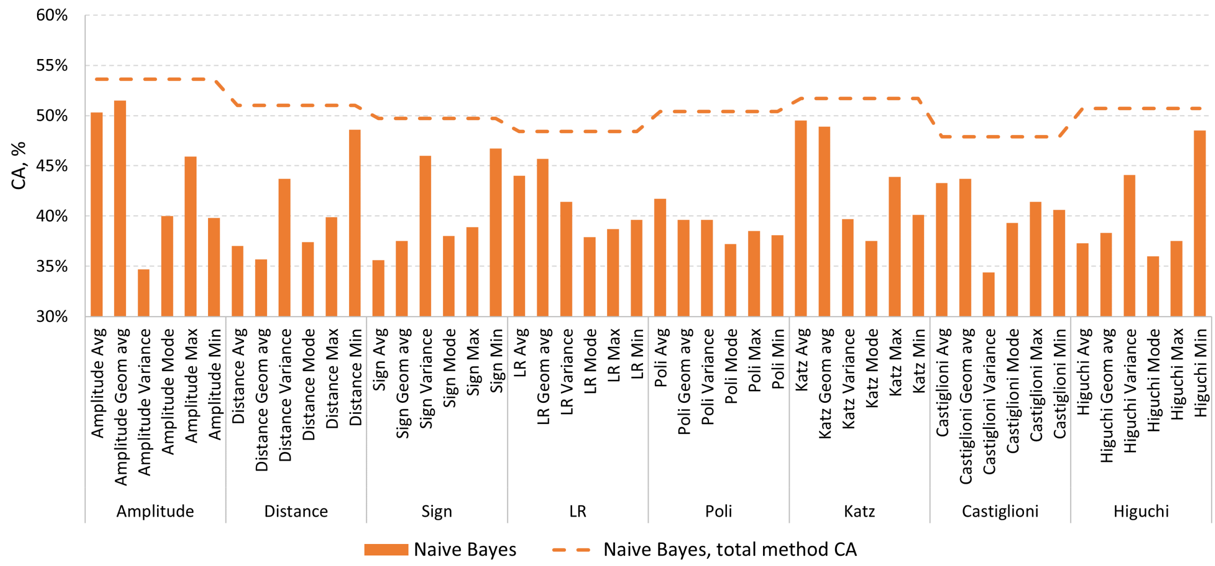


Figure 26. CA by features, Naive Bayes method.

First observation from plotted graphs is a confirmation that each statistical functional performs significantly different in different methods. For example, “avg” feature accomplishes very good results in LR algorithm, while in Distance and Sign algorithms it is completely unreliable. Moreover, we can affirm that any feature alone never reaches better results than a combination of all six. However, the combination of several features (Top 3, Top 5) could result in higher accuracy. Therefore, we performed the following experiment and visualized the outcome (Fig. 27). The experiment was made only with Neural network classifier, as it turned out to be the most accurate in both, methods and parameters, CA rates (4). Also, the trends of each feature performance (Fig. 24, 25, 26) are the same.

Table 4. Avg CA % by different classifiers.

Emotions’ classification	kNN	Neural network	Naive Bayes
avg methods’ CA, %	50.3 %	53.2 %	50.4 %

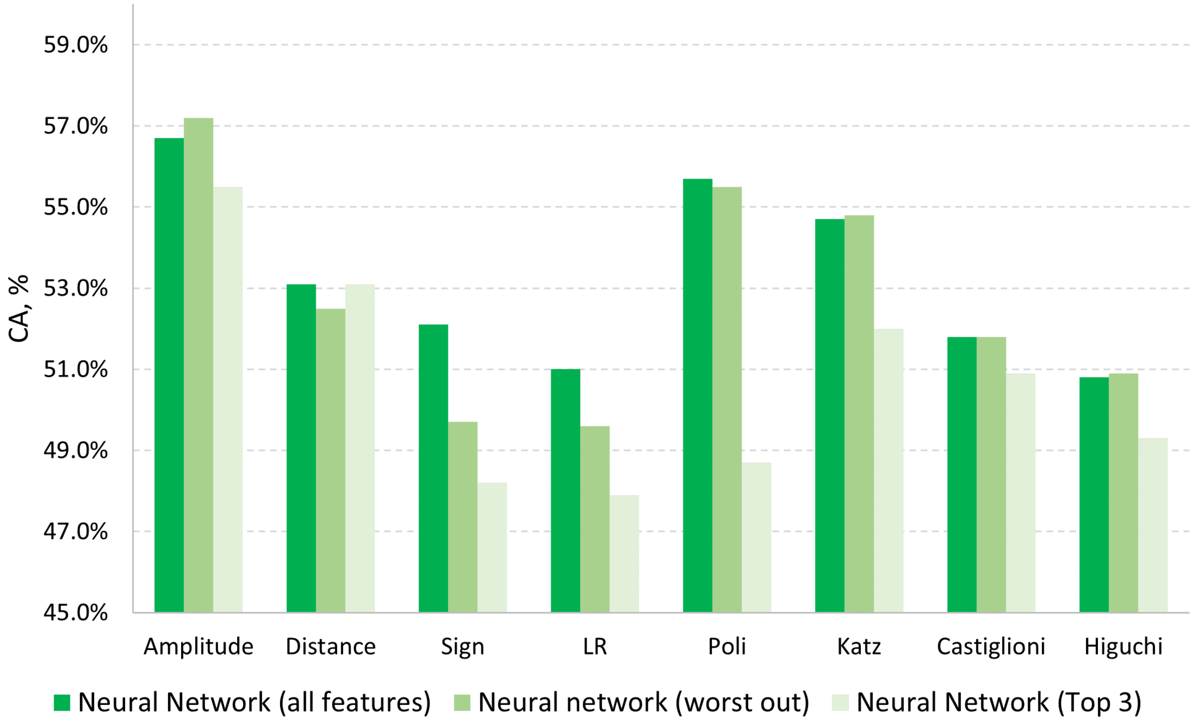


Figure 27. CA by different features' combinations.

In Fig. 27 we can see that the combinations of features did not give very fruitful results: when worst feature was left out, only Amplitude, Katz and Higuchi slightly improved (+0.5%, +0.1% and +0.1% respectively), and all other methods suffered. Taking best 3 parameters even worsened the situation — significant accuracy drop occurred in all methods except Distance. This situation tells us that each feature, even the worst one, positively impacts final results. The general trend is that the more statistical parameters we take, the better results we get. Therefore, additional statistical features could improve our classification results.

While the mix of features did not give any useful improvements, we already know (Fig. 22) that combining different methods produces better outcome. Therefore it is worth to check different permutations of our algorithms. (Figure 28 represents the performance of merged proposed methods: all, Top 4 (Amplitude, Distance, Sign and Poli), Top 3 (Amplitude, Distance and Poli) and Top 2 (Amplitude and Poli). In order to keep the balance in comparison, we also joined Katz, Castiglioni and Higuchi features (marked with "K+C+H").

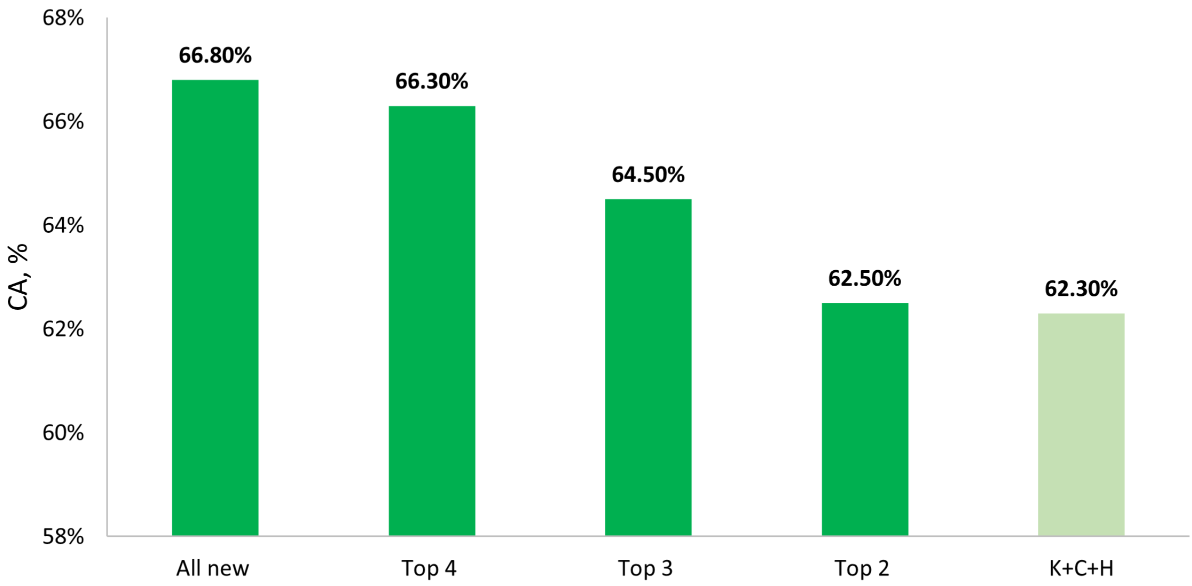


Figure 28. CA by different methods' combinations.

It turned out that not only all features, but in fact all methods play beneficial role in speech emotions' classification task. The worst, LR algorithm, adds +0,5% to total CA rate and proves itself to be at least a bit useful method. Similarly, when Sign features are excluded, results continue worsening. Secondly, observation is that the combination of currently existing and widely approved methods — Katz, Castiglioni and Higuchi — do not catch up with the combination of our proposed methods. Both taken separately and mixed they perform worse.

We also performed an additional analysis (Fig. 30) and checked the accuracy rates' dependency on gender (as our data consists of 10 speakers, 5 men and 5 women). It has come to a light that all classifiers are considerably better in recognizing women's emotions than men's. The accuracy of mix of all new methods improved by 12% and reached 79% when only women were included. Significant impact is seen also in Katz algorithm, which is very good at classifying women and very bad at classifying men. In order to better understand these differences, we extracted confusion matrices (Fig. 29). They show that all emotions are recognized worse, especially sadness. The reasons are unknown, but our guesses are: 1) male speakers portrayed artificial emotions worse 2) men voices were very different and 5 speakers were not enough for a representaive model 3) male voice timbre is strong and amplitude is high in all emotions, thus FD value range is much more narrower for all emotions 4) our models are not comprehensive enough (for example more features are needed) to solve such task with higher accuracy.

		Predicted		
	Women	J	N	S
Actual	J	411	64	25
	N	71	370	59
	S	30	74	396

		Predicted		
	Men	J	N	S
Actual	J	355	55	90
	N	56	314	130
	S	143	145	212

Figure 29. Gender-specific Confusion matrices.

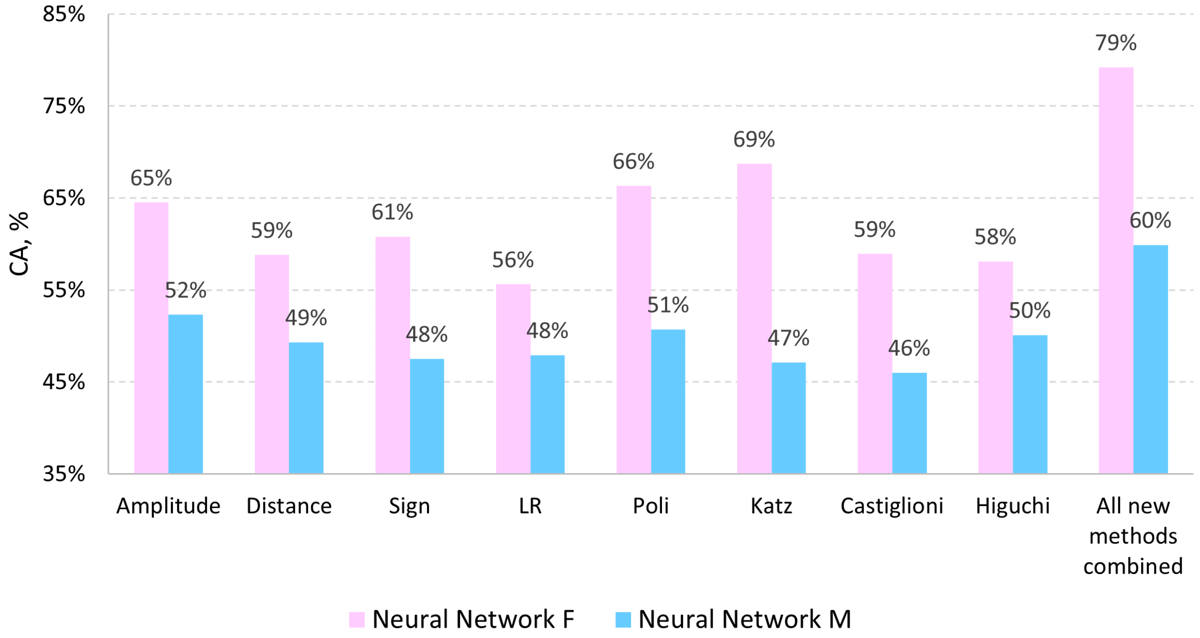


Figure 30. Gender-specific CA rates by different methods.

The final step of our experiment is a speaker-independent emotion classification task. We have applied a “leave one speaker out” principle for testing data separation. In total, ten different cases were analyzed. However, we chose only the best fractal dimension estimation algorithms (All new, Amplitude, Poli and Katz) in order to have less loaded graph and clearer results. The individual CA rates for each speaker are presented in Figure 31. It is important to note that this is the most rigorous testing for the speech analysis-based classification tasks, as the testing is based on unknown data and worse results should be expected.

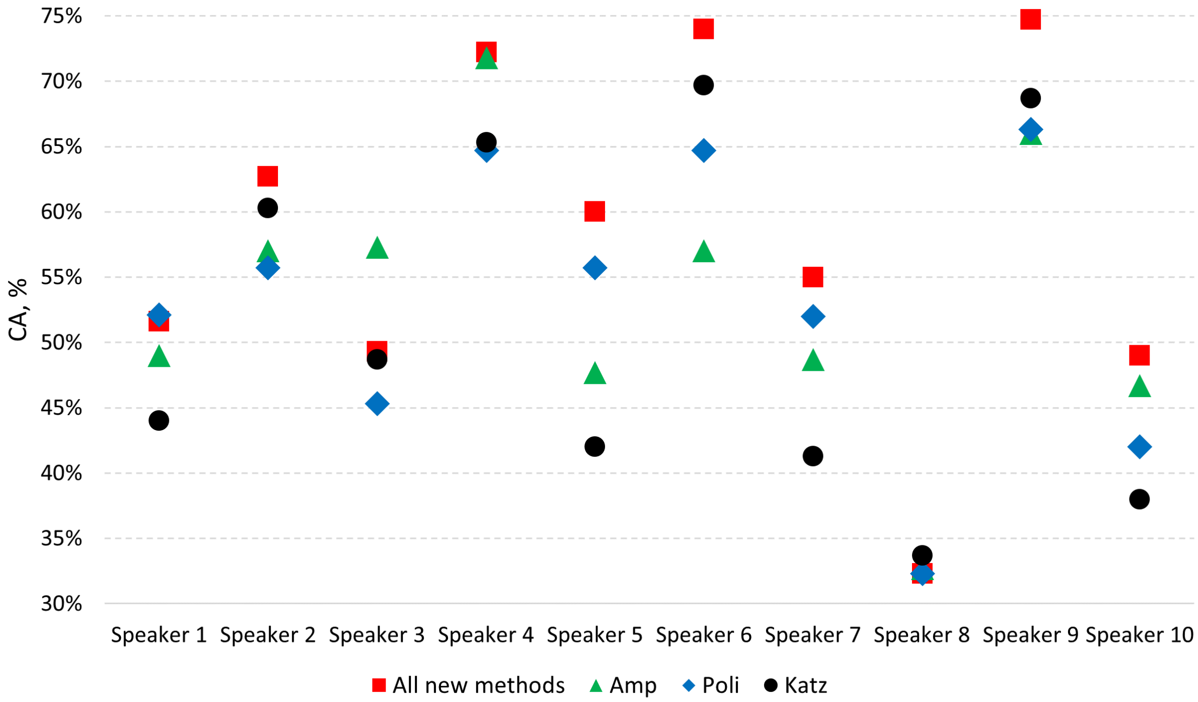


Figure 31. Speaker-specific CA rates by different methods.

In 7 out of 10 cases all new methods performed the best and reached the highest accuracy rates, which means that they are more effective than Katz. Also, Amplitude and Poli algorithms alone are able to compete with Katz algorithm and in some cases provide better results. Katz performed well on female classification (for example Speaker 2, Speaker 6, Speaker 9), but was unreliable at male emotions' recognition. Speaker 8 results differed significantly and could be considered as an outlying. On the other hand, this outlier together with the general differences seen between speakers demonstrate that each individual has its own specific voice and speaking manners. Therefore, it is a very complex task to build a model, which would be very accurate in classifying emotions for all speakers. From the graph we also see that CA rates strongly depend on both the speaker and the method, as the values are vastly scattered. To conclude, we can state, that the proposed methods are neither less reliable, nor less useful than some already existing and acknowledged, and widely used methods.

5 Summary

5.1 Results

In this paper we managed:

- To introduce fractals and fractal dimension together with its estimation algorithms as well as use cases;
- To create a proposal of 5 new different fractal dimension estimation methods. We called them “Amplitude”, “Poli”, “Sign”, “LR” and “Distance”;
- To analyze theoretical properties (fractal dimensions of extreme cases, harmonic function, Weierstrass sine function) of new methods;
- To compare theoretical reliability of new versus few already existing methods (Katz, Castiglioni, Higuchi);
- To make a small experiment on different sounds recognition. New methods were able to distinguish some groups of sounds and managed visually separate vowels from consonants;
- To carry out a comprehensive speech emotion recognition experiment (Joy, Neutral and Sadness):
  - We calculated the fractal dimension of each 20 ms in every recording using 5 new and 3 old methods;
  - From obtained FD sequences we calculated statistical functionals, such as average, variance, etc;
  - We loaded the data to three classifiers (Neural network, kNN and Naive Bayes) and applied 10 fold cross-validation scheme;
  - We experimented with different features and methods combinations;
  - We performed a speaker-independent classification task;
  - We compared achieved accuracy rates of new methods versus old methods in a very detailed manner.

Our main achievement is that the proposed methods were able to both, theoretically and practically, compete or even outperform Katz, Castiglioni and Higuchi methods.

5.2 Conclusions

Main conclusions:

1. Fractal modeling of speech signals is a valuable and useful tool;

2. We have proposed 5 new methods; in theoretical applications the most successful methods were Amplitude, Sign and LR, in practical experiment - Amplitude and Poli. In most cases new methods' performance was better than Katz, Castiglioni and Higuchi;

3. A lot of methods might perform well in theoretical tasks, but have disadvantages while working with real-life data, and vice versa. In other words, good theoretical results do not guarantee good practical results, therefore it is always important to carry out a practical experiment.

Some additional conclusions, which we noticed from emotion recognition task:

- Combined methods can result in better accuracy rates;
- Each feature input to a model is different, but the more features are included, the better the outcome is;
- Artificial neural network based classifier enabled us to obtain the best speech emotion identification results.

5.3 Discussions

Fractal modeling of speech signals is a wide area and we believe that even more detailed analysis should be made. Our proposals for future research:

- More currently existing fractal dimensions could be considered and compared (in both theoretical and practical parts);
- More variations of proposed methods could be included (for example all methods with and without logarithms could be considered, interval split to Amplitude method could be applied);
- Deeper theoretical examination should be made: more curves with known fractal dimension could be incorporated (for example Weierstrass cosine function, Takagi function), proposed methods' convergence to true fractal dimension could be analyzed;
- During the experiment, several improvements could help:
  - Bigger data set;
  - Overlapping intervals could be used for fractal dimension calculation;
  - More statistical features extracted (for example quantiles, central moments, etc);
  - More advanced classifiers (for example optimal parameters could be found for neural network);
  - Applied feature selection techniques.

These improvements would increase the reliability of the research and, we believe, that the new methods could then be suggested for a wider audience and be used all across the world.

6 References

[1] L. Debnath, "A brief historical introduction to fractals and fractal geometry," *International Journal of Mathematical Education in Science and Technology*, pp. 29–50, 2006.

[2] B. Mandelbrot, *The Fractal Geometry of Nature.* 1983.

[3] M. E. Gaddis and M. J. Zyda, *The Fractal Geometry of Nature: Its Mathematical Basis and Application to Computer Graphics.* 1986.

[4] K. Falconer, *Fractal Geometry. Mathematical Foundations and Applications.* Wiley, 1986.

[5] P. Maragos, "Measuring the fractal dimension of signals: Morphological covers and iterative optimization," *IEEE Transactions on Signal Processing, vol. 41*, pp. 108–121, 1993.

[6] K. Falconer, *Fractal Geometry: Mathematical Foundations and Applications.* 1990.

[7] S. Fekkai, *Fractal based speech recognition and synthesis.* 2002.

[8] M. Fernandez-Martinez, "Theoretical properties of fractal dimensions for fractal structures," *Discrete and continuous dynamical system series, vol. 8, number 6*, pp. 1113–1128, 2015.

[9] X. Huang, A. Acero, and H.-W. Hon, *Spoken language processing.* 2001.

[10] A. Juozapavičius and T. Meškauskas, *Vaizdų ir signalų analizė ir apdorojimas.* 2011.

[11] R. S. Holambe and M. S. Deshpande, *Advances in Non-Linear Modeling for Speech Processing.* 2012.

[12] D. J. Tritton, *Physical Fluid Dynamics.* 1988.

[13] L. Huang, "The sound of wind: Vortex shedding and how it works." `https://steemit.com/science/@frostfractal/the-sound-of-wind-vortex-shedding-and-how-it-works`.

[14] Apple, "Apple site." `https://www.apple.com/homepod/`.

[15] B. Klinkenberg, "A review of methods used to determine the fractal dimension of linear features," *Mathematical Geology, vol. 26*, pp. 23–46, 1994.

[16] N. Rajković, B. Krstonosic, and N. Milošević, "Box-counting method of 2d neuronal image: Method modification and quantitative analysis demonstrated on images from the monkey and human brain," *Computational and Mathematical Methods in Medicine*, 2017.

[17] M. Y. Marusina and E. A. Karaseva, "Application of the box-counting method for the evaluation of medical images," *2019 International Conference Quality Management, Transport and Information Security, Information Technologies*, 2019.

[18] M. J. Katz, "Fractals and the analysis of waveforms," *Computers in Biology and Medicine, vol. 18*, pp. 145–156, 1988.

[19] D. R. Jevtic, "Application of katz algorithm for fractal dimension in analysis of room impulse response," *Telecommunications Forum (TELFOR) Proceedings of Papers*, pp. 1063–1066, 2011.

[20] B. S. Raghavendra and D. N. Dutt, "A note on fractal dimensions of biomedical waveforms," *Computers in Biology and Medicine*, pp. 1006–1012, 2009.

[21] C. Goh, B. Hamadicharef, G. Henderson, and E. Ifeachor, "Comparison of fractal dimension algorithms for the computation of eeg biomarkers for dementia," *CIMED2005 Proceedings*, pp. 463–471, 2005.

[22] H. Eguiraun, K. Lopez Ipina, and I. Martinez, "Application of entropy and fractal dimension analyses to the pattern recognition of contaminated fish responses in aquaculture," *Entropy, vol. 16*, 2014.

[23] G. Tamulevicius, R. Karbauskaite, and G. Dzemyda, "Speech emotion classification using fractal dimension-based features," *Nonlinear Analysis: Modelling and Control, vol. 24*, 2019.

[24] P. Castiglioni, "What is wrong in Katz's method? Comments on: "A note on fractal dimensions of biomedical waveforms"," *Computers in Biology and Medicine, vol. 18*, pp. 950–952, 2010.

[25] K. Lopez-de Ipiña, J. Solé-Casals, H. E. Martinez, and J. B. Alonso, "Feature selection for spontaneous speech analysis to aid in Alzheimer's disease diagnosis: A fractal dimension approach," *Computer Speech and Language , vol. 30*, 2014.

[26] K. Lopez-de Ipiña, A. Bergareche, P. de la Riva, M. Faundez-Zanuy, P. M. Calvo, J. Roure, and E. Sesa-Nogueras, "Analysis of non-invasive writing signals, applied to Essential Tremor: A nonlinear approach," *Journal of Applied Logic*, 2015.

[27] T. Higuchi, "Approach to an irregular time series on the basis of the fractal theory," *Physica D: Nonlinear Phenomena, vol. 31*, pp. 277–283, 1988.

[28] A. Accardo, M. Affinito, M. Carrozzi, and F. Bouquet, "Use of the fractal dimension for the analysis of electroencephalographic time series," *Biological Cybernetics, vol. 77*, pp. 339–350, 1997.

[29] P. Paramanathan and R. Uthayakumar, "An algorithm for computing the fractal dimension of waveforms," *ScienceDirect, vol. 195*, pp. 598–603, 2008.

[30] E. Olejarczyk, "Application of fractal dimension method of functional MRI time-series to limbic dysregulation in anxiety study," *Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 3408–3410, 2007.

[31] S. Kesić and S. Z. Spasić, "Application of Higuchi's fractal dimension from basic to clinical neurophysiology: A review," *Computer Methods and Programs in Biomedicine, vol. 133*, pp. 55–70, 2016.

[32] V. O. Nams, "Improving accuracy and precision in estimating fractal dimension of animal movement paths," *Acta Biotheoretica, vol. 54*, pp. 1–11, 2005.

[33] R. Andrle, "Estimating fractal dimension with the divider method in geomorphology," *Geomorphology, vol. 5*, pp. 131–141, 1992.

[34] M. S. Raimundo and Jun Okamoto Jr, "Application of Hurst Exponent (H) and the R/S Analysis in the Classification of FOREX Securities," *International Journal of Modeling and Optimization, vol. 8*, pp. 116–124, 2018.

[35] Y.-Z. Wang, B. Li, R.-Q. Wang, J. Su, and X.-X. Rong, "Application of the Hurst exponent in ecology," *Computers and Mathematics with Applications, vol. 61*, pp. 2129–2131, 2011.

[36] L. Zao, R. Coelho, and P. Flandrin, "Speech Enhancement with EMD and Hurst-Based Mode Selection," *IEEE Transactions on Audio, Speech, and Language Processing, vol. 22*, pp. 899–911, 2014.

[37] A. Petrosian, "Kolmogorov Complexity of Finite Sequences and Recognition of Different Preictal EEG Patterns," *IEEE Symposium on Computer Based Medical Systems*, pp. 212–217, 1995.

[38] C. Sevcik, "A procedure to Estimate the Fractal Dimension of Waveforms," *Complexity International, vol. 5*, pp. 1–9, 1998.

[39] L. T. Wah, *The Fractal Dimension of the Weierstrass Type Functions.* 1998.

[40] Orange Data Mining, "Orange platform." `https://orange3.readthedocs.io/projects/orange-visual-programming/en/latest/index.html`.