# Nonlinear Assessment of the Professional Voice "Pleasantness"

Horia Nicolai Teodorescu[1,2]
Monica Feraru[1]
Diana Trandabat[3]

[1]Technical University of Iasi, Center CERFS, Iasi, Romania
[2]Romanian Academy
[3]Institute for Theoretical Informatics of the Romanian Academy

## Overview

- Introduction
- What makes a professional voice agreeable?
- Methodology
- Results
- Conclusions

## Introduction

- Speech is a rich and subtle communication manner
- It transfers:
  - linguistic data
  - speaker's personality
  - emotional states
- Professional voice = a voice used in professional relations

  (Example: singers, drama players, lawyers professors, politicians, media speakers..)

## Voice pleasantness

- The pleasantness of the voice is determined by: ***prosody*** and ***voice quality***
  - Prosody = meta-language that carries information on the speaker and on his/her relationship to the overall context
  - Voice quality = identity of the voice through its vocal timbre

## What makes a professional voice agreeable?

- "pleasant" and "natural" voice depends on the degree of shimmer and jitter
  - A higher shimmer and jitter degree produces an unpleasant voice
  - A lower shimmer and jitter degree produces an agreeable voice
- The differences between natural pleasant and unpleasant voices are used to assess the synthesized voices quality

## What makes a professional voice agreeable?

- Relevant parameters for the prosodic information and the expression of emotions in speech:
  - the mean and the variance of the fundamental frequency within an utterance
  - the mean and the variance of the intensity
  - the breaks frequency
  - the debit
  - modification of the degree of jitter and shimmer
- Ex: the anger, the fear and the intense joy:
  - Augmentation of F0
  - Augmentation of the intensity
  - Decrease of the duration of specific segments

## What makes a professional voice agreeable?

- Voice shimmer and jitter may significantly degrade the subjective voice quality.
- The degree of shimmer and jitter helps to identify emotions like fear, sadness and anger.
- Therefore, we need to analyze and assess the degree of shimmer and jitter in voice signals to determine the emotion.

## Methodology

- The data:
  - recordings of 20 students, after signing an informed consent
  - we selected a segment of 100ms to 500ms of the registration for analysis.
  - the original .wav files are transformed into a numeric, .txt file.

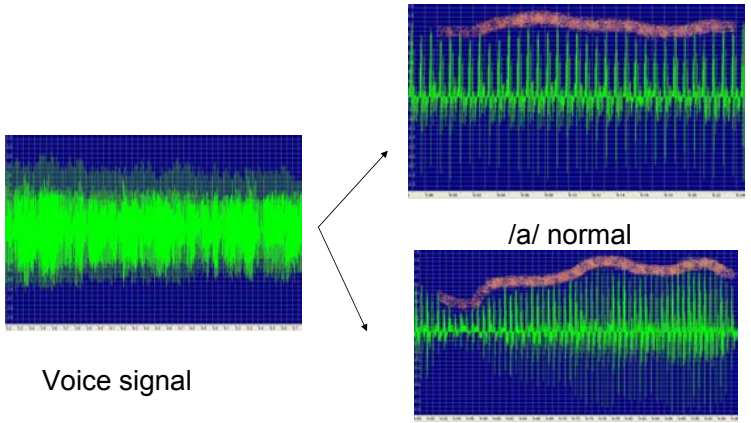
# Methodology

- Analysis:
  - segments containing shimmer or jitter were selected from the sampled vocal signal using visual and acoustical inspection
  - the speech signal segment is analyzed using the fractal dimensions, using a free software, the CDA analysis package:
    - the Lyapunov exponent
    - the correlation dimension
    - the capacity dimension

# Methodology

| | |
|---|---|
| The correlation dimension | $D_c = \lim_{\varepsilon \to 0} \frac{\ln \sum_{i}^{N(\varepsilon)} P_i^2}{\ln \varepsilon}$ |
| The Lyapunov coefficients | |
| The capacity dimension | |

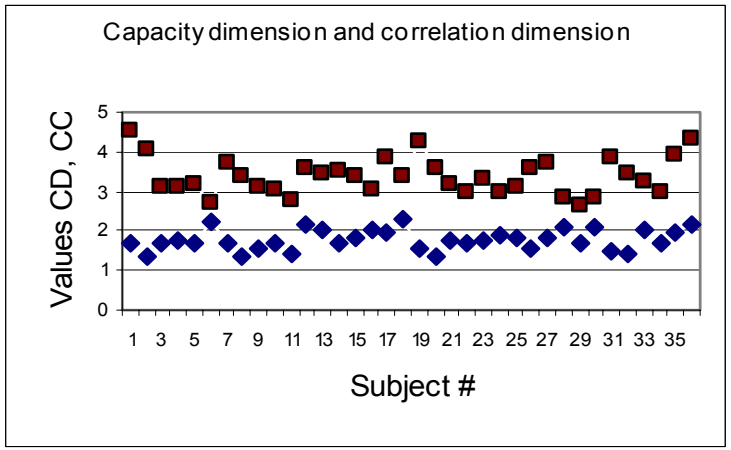
# Methodology



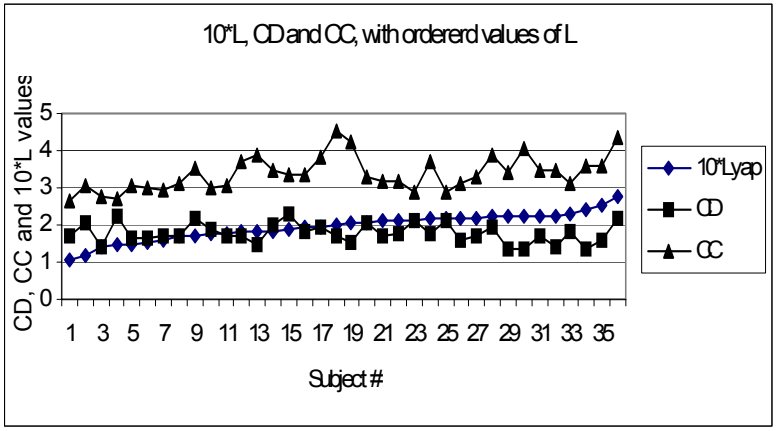

# Results

- We obtained good discrimination results for eight subjects and acceptable discrimination results for six (for the /e/ vowel).
- We have noticed that the capacity dimension provides complementary information to the Lyapunov exponents and improves discrimination.

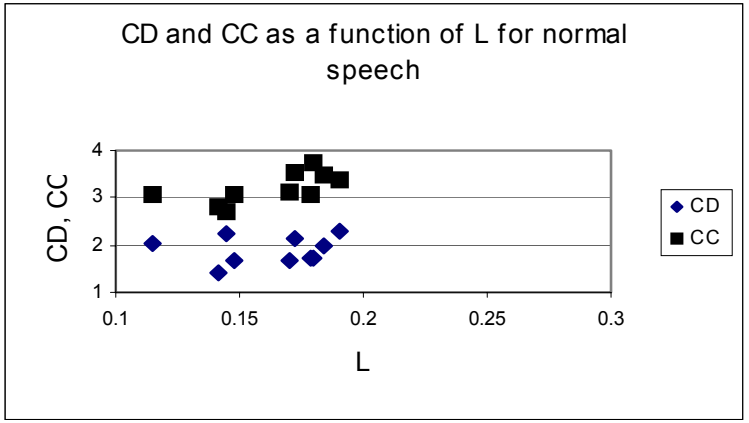
# Results



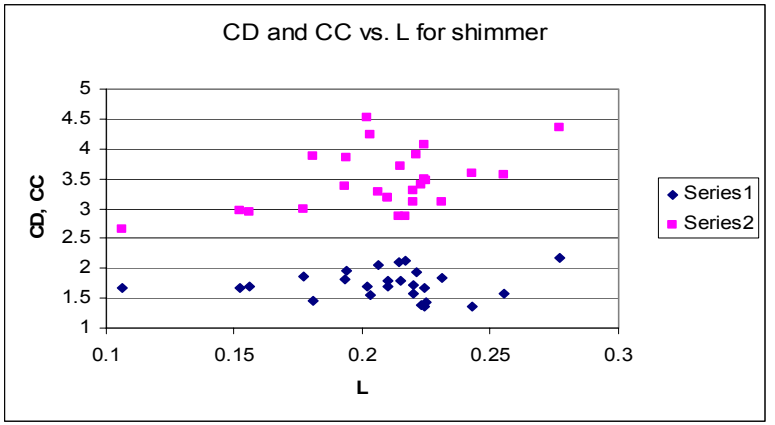

# Results




# Results




# Results



## Results

| | |
|---|---|
| Correlation coefficient between L and CD for normal signal | 0.42 |
| Correlation coefficient between L and CD for tremor | 0.19 |
| Correlation coefficient between L and CC for normal signal | 0.53 |
| Correlation coefficient between L and CC for tremor | 0.57 |

## Conclusions

- The capacity dimension is an informative parameter and gives complementary results to the Lyapunov exponent.
- For the tremor voice signal, we notice a modest increase.
- Shimmer and jitter in natural voices relate to nonlinear processes

## Conclusions

- Shimmer and jitter analysis is useful to assess the quality of the natural voices and to increase the naturalness of the artificial voices.
- The use of shimmer and jitter analysis and synthesis may help endowing the machines with “emotional intelligence” (e.g. emotional expressivity, ability to recognize emotion from vocal signals and to simulate emotion when confronted with a specific situation).

# Thank you