Digital Signal Processing

Laboratory

Final Practice

Vowel Classifier

miguelangel.garcia.izquierdo@upm.es

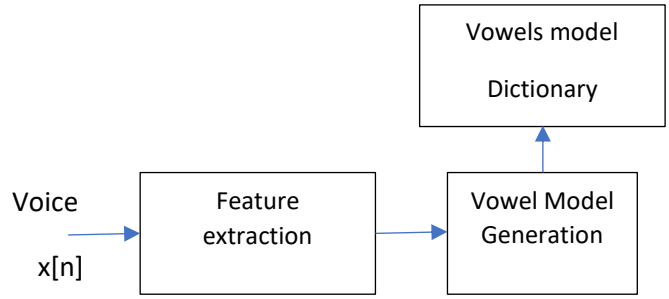
B-408

1 Introduction

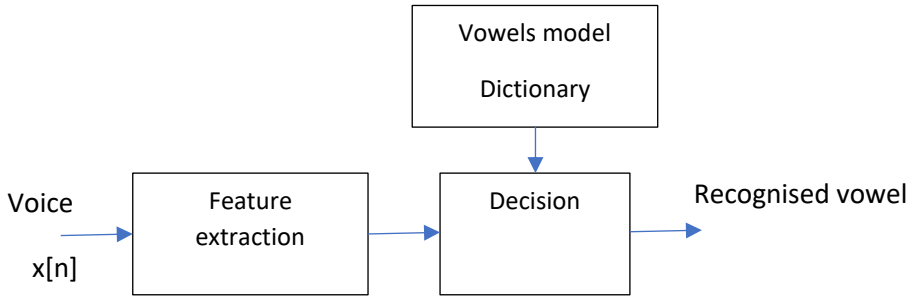
Speech is one of the most useful and complex ways of human communication. The information carried by the voice signal can be an idea, characteristics of a speaker, a sequence of allophones, etc.

This practice deals with the recognition of some isolated allophones: the vowels. For this, it will be necessary to have a database with enough repetitions of each of them, digitized at 8000 Hz and pre-emphasized with the filter $H(z) = 1 - 0.95z^{-1}$.

The recognizer system will have two distinct stages: training and recognizer. Both will have the feature extraction block in common.



Training stage



Recognition stage

1.1 Extraction of the Feature Vector

The signal $x[\,n\,]$, after being pre-emphasized, is windowed with a 256 samples (32 ms) Hamming window, $w[\,n\,]$ with 50% overlapping (128 samples, 16ms), obtaining $x_w[\,n\,] = x[\,n\,] \cdot w[\,n\,]$. From each of these widowed subsequences, the cepstrum coefficients are obtained:

$$c_i^k\,[\,n\,] = FFT^{-1}\,\{\log[\,|\,FFT(\,x_w\,[\,n\,]\,)\,|\,]\,\} \quad 0 \le n \le 255$$

Indexes refer to the *i-th* window of the vowel *k*.

Only the first 10 cepstrum coefficients are selected for obtaining the feature vector of the windowed subsequence.

$$\mathbf{v}_i^k = \{\, c_i^k\,[\,0\,], c_i^k\,[\,1\,], c_i^k\,[\,2\,], c_i^k\,[\,3\,], c_i^k\,[\,4\,], c_i^k\,[\,5\,], c_i^k\,[\,6\,], c_i^k\,[\,7\,], c_i^k\,[\,8\,], c_i^k\,[\,9\,]\,\}$$

This process is repeated for all the windowed subsequences of the vowel and for all the vowels.

1.2 Generation of the vowel model

For each vowel, *k*, all vectors, *i*, are used to generate a codebook of 4 centroids using the LBG algorithm described in [2] and [3] that must be implemented. The obtained codebooks for all the vowels will form the vowels model dictionary.

1.3 Comparison Algorithm

During the recognition stage, for each feature vector and for each codebook *k*, the Euclidean distance is computed.

1.4 Decision

The index k of the codebook with the minimum sum of the Euclidean distances is the recognized vowel k.

1.5 Objective of this work

The objective of this work is the programming in MATLAB of the training and recognition blocks described above. You can decide to make the recognizer speaker-independent by using many recordings of vowel repetitions from multiple speakers, or speaker-dependent if you only use one.

2 Deliverables

2.1 Methodology

Groups

This work will be done in groups of 4 students.

Elaboration

This work will be divided into two phases that in turn translate into two deliverables:

Phase 1. First deliverable: Analysis and design of the proposed work

In this phase, an analysis of the problem and the methodology to be applied will be made. The result will be a memory that, at least, must include:

• Previous study. Description of the two concepts on which this practice is based: cepstrum and LGB algorithm. Summary of the most important aspects.

• Workplan. Methodological analysis of the project: This part must include a decomposition into tasks, and the assignment of tasks to each of the group's components.

• Software design: Determination of the structure of the program and of the prototypes of the functions to be implemented. (not its implementation). In this stage, the prototypes of all the functions that are going to be programmed must be presented. As a suggestion, the LGB function is presented.

```
function codeBook = LGB( vowelCepstrums )
% Finds the model containing the 4 centroids for the vowel cepstums that
is passed as parameter.
% Inputs:  vowelCeptrums. Nx10 matrix tha contains all the cepstrums of
one vowel.
% Outputs: codeBook. 10x4 matrix that contains the 4 centroids
codebook of the vowel.
```

Phase 2. Second deliverable: Implementation

In this phase, the code development, integration, and testing of the functions that make up the project will be carried out. The results will be the second deliverable:

• Memory with the description of the coding of the solution, the tests and results obtained.

• MATLAB codes and data files that make up the solution.

2.2 Evaluation

The evaluation of the works will be based on two elements:

• Presentation of the work, which will be done collectively with an approximate duration of 10 minutes. This presentation must include a demonstration of the results obtained. Next, the teacher will ask questions related to any aspect of the work, which will be formulated individually to each of the components of the group after the presentation. (50% of the final work grade)

• Individual exam on aspects related to the work carried out (50% of the grade of the final work).

3 Calendar

Deadlines:

30/09 Groups configuration.

18/11 Presential first deliverable delivery and mandatory tutorial. Must be programed with the teacher earlier.

16/12 Presential second deliverable delivery and project presentation.

4 Bibliography

[1] Sadoaki Furui, Digital Speech Processing, Synthesis and Recognition, Mercel Dekker

[2] Xuedong Huang and Alex Acero, Spoken Language Processing, Prentice Hall

[3] Y. Linde, A. Buzo and R.M. Gray, “An Algorithm for Vector Quantizer Design”, IEEE Trans. Communications, pp. 84-95, Jan.1980

[4] Discrete-time Signal Processing, Alan V. Oppenheim and Ronald W. Schafer, Prentice Hall