# AUTOMATIC DETECTION AND ANALYSIS OF DYSPHONIA

**Kamil Ekštein**

*Laboratory of Intelligent Communication Systems,*
*Dept. of Computer Science and Engineering,*
*Faculty of Applied Sciences, University of West Bohemia,*
*Univerzitní 22, 306 14 Plzeň, Czech Republic*
*kekstein@kiv.zcu.cz*

Abstract: This article deals with automatic methods of dysphonia diagnostics and analysis. A distributed automatic diagnostic and analysis system **DAn** (Dysphonia Analyser) is described. A detailed description of the implemented diagnostic and analysis methods is given together with some relevant physiological particulars of dysphonia.
The DAn project was started as a joint research activity between Laboratory of Intelligent Communication Systems and Teaching Hospital in Plzeň. The motive for the research and development of the system was the need to objectively diagnose patients with voice disorders. Before the DAn system was launched the used diagnostic method had been repeated subjective expert auscultation performed by several speech therapists (frequently only one due to capacity reasons) without any technical equipment.
Also the measuring setup was refined during the research stage of the project—a special, highly sensitive microphone with linear transmission characteristics was chosen and several experiments were performed to figure out the best microphone position and distance in order to maximise the possible dysphonia traits in the recording for both automatic analysis and expert listening.
The DAn system enables the therapist to either diagnose the patient in a fully automatic mode or examine him/her thoroughly in an expert-assisting mode and then make the diagnostic conclusions. In the fully automatic mode the system analyses the recording and informs the therapist about dysphonia symptoms it possibly found. The expert-assisting mode helps the therapist to decide him/herself. In this mode the system amplifies and filters the recorded signal and acoustically emphasises the spectral regions that are important for expert diagnose.

Keywords: speech signal processing, speech signal analysis, dysphonia detection, dysphonia diagnosis

## 1 INTRODUCTION

Dysphonia is a voice disorder which is characterised mainly by presence of noise components in spectra of vocals (like e.g. long [a:] vowel which is usually used for testing). The dysphonia is not only a social disease: It is, of course, a cardinal problem for any speech professional. However, it is far more important from the medical point of view as an indicator of structural changes in the vocal tract caused by e.g. a cancer. Therefore it is very important to properly diagnose the dysphonia in its early stages so that a treatment can be started in due time.

Considering the above said, the detection and analysis methods are based primarily on a high-resolution spectral analysis and digital signal filtering. The final diagnosis decision (i.e. classification into 2 disjunct classes—dysphonia and non-dysphonia) is taken either by an expert (or a group of experts) or an artificial neural network depending on the working mode (see 3).

The **DAn** system is generally designed as a client-server application composed of several more or less stand-alone modules. The **LDAn** application (see Fig. 1.) serves both as a client for accessing the remote database and as a working environment for the expert who can use it to analyse/diagnose patient samples in either a fully automatic or an assisted mode.

For those experts and facilities whose computers are not powerful enough or have not enough storage capacity, there is a *web-based (HTTP) client*. The web-based client enables the user to (i) upload the recorded samples to the central recording database, (ii) browse the stored samples in order to find e.g. similar ones, and (iii) to have the sample analysed by a *server-operated analysis engine*.

The image below shows the system architecture:

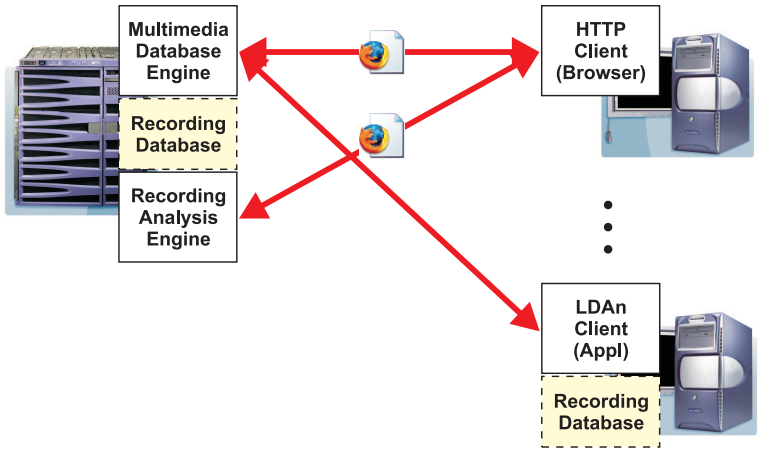


Fig. 1. Client-server architecture of the distributed **csDAn** system.

### *2.1 Server*

The main purpose of the **DAn Project** server is to manage the *recording database* via a multimedia database engine—both analysed/diagnosed and 'raw' recordings are stored. Further it provides an access to the automatic analysis/diagnostic engine (which is also a part of the **LDAn** client application) and ensures the distribution of the up-to-date setup file for the neural network of the client applications—in other words it manages the training of the analysis engine for the clients.

The *recording analysis engine* is entirely the same as in the FAD mode of the **LDAn** client application (see 3.1). However, the server benefits from its power and performs the training of the neural network using all possible data from the database[1] during nights (or time when the server is idle).

### *2.2 Clients*

There are two types of clients in the **DAn Project** framework: (i) *web-based (HTTP) clients* and (ii) **LDAn** applications running on experts' desktop computers.

The simple, web-based client enables the user to browse the database of recordings and view

[1] The training process takes typically tens of hours and thus it is unsuitable for desktop computers of the clients.

the diagnoses and remarks associated to the sample files. This mode is designed particularly for medical students and those ORL practitioners that do not want to have the whole system (microphone, headphones, software) installed in their surgeries but want only to compare the samples with e.g. their own auscultation results.

The **LDAn** client application can work in two different modes (see 3): In the expert-assisting mode it helps the expert to record, analyse and diagnose the sample via various signal processing techniques and an integrated working environment. In the fully automatic mode the software tries to analyse the sample, detect the essential features and provide the used with a diagnosis. Moreover the **LDAn** client application can be used to browse the recording database on the **DAn Project** server too and besides it can manage a local recording database created by the user.

## 3 FUNCTION DESCRIPTION

### *3.1 Fully Automatic Diagnostics Mode*

The **fully automatic diagnostic** (FAD) mode is designed to provide the otorhinolaryngologist/speech therapist with a diagnosis of the analysed sample. This FAD mode is based on the performance of a *multi-layer perceptron network* (MLP) which implies that accurate results can be obtained only in the case that the neural network is trained with enough data. Thus, in the early stages of the project, this mode cannot entirely replace the expert auscultation. However, after gathering a large database of correctly diagnosed (classified from the neural network point of view) samples for the network training, it is estimated to be able to reach 95 % (or better) accuracy. Such a number vastly outperforms majority of common experts who are able to diagnose the dysphonia at some 50 - 60 %.

The automatic diagnostics procedure description follows.

**Diagnostics procedure:**
A patient that is diagnosed is instructed to utter a long vowel [a:] (like e.g. in 'half'). His/her voice is recorded in a small, quiet, natural reverb room using a high-quality condenser microphone (we used the AKG C4000B) and digitised at 44.1 kHz/16 bits per sample (CD quality). Subsequently the recorded signal is processed in the following steps:

1. **Pitch synchronization** — in order to gain the best possible spectral resolution without effects of boundary cuts, the analysed sample is divided into frames in such a way that the phase angle of the speech signal in each frame is zero. It is achieved by cutting the signal at the positions where the sample value is zero or where the (interpolated) zero value lies between two consecutive samples.

2. **Windowing** (93 msec, i.e. 4096 samples at 44.1 kHz) — unlike in speech recognition, here it is necessary to analyse longer sections of the signal so that the frequency resolution is better (here approx. 11 Hz per value). The window is *rectangular*, i.e. no weighting function is applied because it is not necessary thanks to the pitch synchronisation.

3. **Preemphasis** ($a$ = 0.99) — a preprocessing technique which emphasises spectral components with increasing frequency (because the dysphonia proves as improper noises at high frequencies of vowel spectra). Implemented as a digital filter, namely *first-order FIR* filter (see [1]) of which recurrent formula is $y[i] = x[i] - a \cdot x[i-1]$, where $y[i]$ is

the preemphasised output signal, $x[i]$ is the original signal, $i \in \langle 1, N \rangle$ is a sample index in the N samples long frame, and $a$ is the *preemphasis coefficient*.

The image below depicts the impact of preemphasis on the signal spectrum:

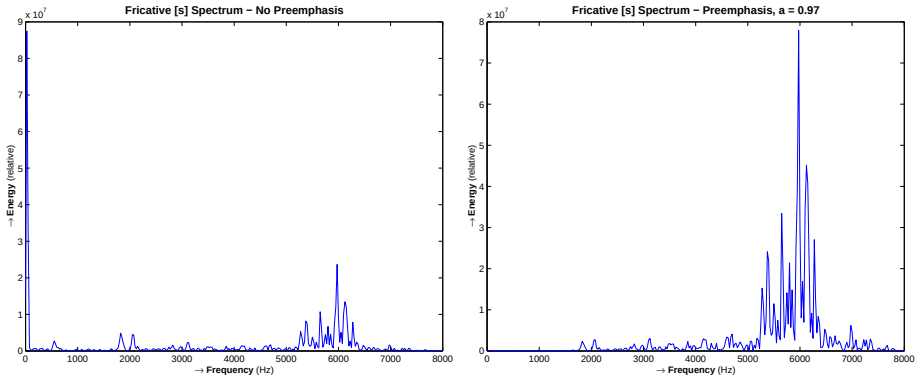


Fig. 2. Power spectrum of a fricative speech sound signal without preemphasis (left) and preemphasised with $a = 0.97$ (right).

4. **Normalisation** — transforms the signal so that its amplitude values do not exceed a given range. The dysphonia detection method bases the classification on the energies of the signal at certain frequencies and thus it is necessary to ensure roughly the same average energy values over the whole frame for all the analysed samples.

5. **Frequency analysis** (via FFT) — the *Fast Fourier Transform* converts amplitude values in the temporal domain into energies in the frequency domain, i.e. into a *power spectrum*. As a result a vector of 2048 energies in approx. 11 Hz-wide bands is obtained.

6. **Neural network processing** (MLP) — the three layer perceptron network takes the 2048-point power spectrum as its input and computes the excitation of two output neurons. One neuron signalises the presence of the dysphonia symptoms while the other the opposite.

The MLP network setup file (synaptic weights, biases, etc.) can be downloaded from the **DAn Project** server. The MLP setup file is computed whenever the MLP of the server analysis engine is retrained with a new data (if available). If the client downloads the retrained setup file then the client MLP is entirely the same as that on the server and thus the analysis engine provides the same results.

### *3.2 Expert-assisting Diagnostics Mode*

The **expert-assisting diagnostics** (EAD) mode is designed to help the otorhinolaryngologist/speech therapist to analyse and diagnose the recorded sample. This mode is intended to be used in the early stages of the project to gather the database of the correctly diagnosed samples for the FAD mode training. However, it is not the only purpose of the mode. As the developed system is designed for educational purposes too, the EAD mode can be used to teach medicine students how to analyse the samples, what to look for, and how to designate the diagnose.

In the EAD mode it is possible to play back the whole sample or its parts. Each part marked for the playback may be looped. The software also allows the user (i) to emphasise certain frequency areas using a graphic equalizer, (ii) to view the signal as a spectrogram, and (iii) to filter the signal using a bandpass filter (or set of such filters). The expert performs the auscultation using high-quality headphones (e.g. AKG K271, Sennheiser HD250, or Sennheiser HD280).

The EAD mode software can be used as a browser through the database of the recorded samples. When connected as a client to the **DAn Project** server, the user can go through the available diagnosed samples and can also upload the analysed sample to the database — with or without a diagnose. If the sample is uploaded without a diagnose, it can be filed into an *assessing process*—any of the associated/co-operating experts can put his/her diagnose (or any related opinion) onto a pinboard connected to the given sample.

## 4 CONCLUSION

The system is currently under the development and works only in a restricted extent. The analysis method performance is validated in discussions with otorhinolaryngology experts. Also the data for the recording database are collected so that the MLP can be re-trained with a large enough training set.

Moreover a serious legal problem appeared during the testing operation of the whole system: According to Czech law the medical data (including the recordings) must be treated as confidential and as such they cannot be uploaded onto a publicly available server. Thus the **DAn Project** server security must be legally solved before the system can be started—unfortunately the law is not easily comprehensible what has precluded the launching of the system yet.

## REFERENCES

[1] Vích, R. and Smékal, Z. *Číslicové filtry (Digital Filters)*. Academia, Prague, 2000. ISBN 80-200-0761-X.