NannyRecord

1. BASIC INFORMATION

1.1. Tool Name

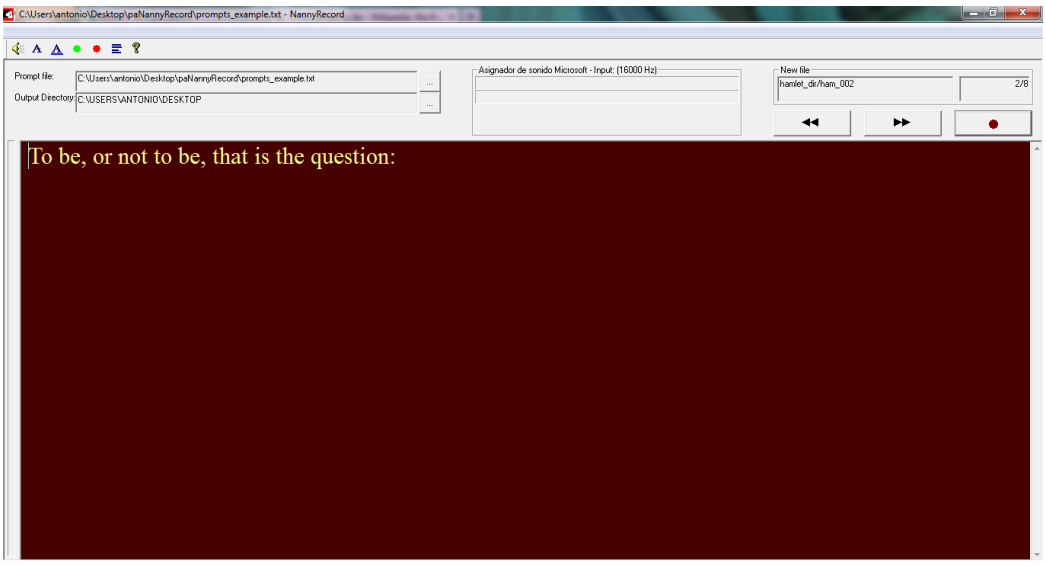
NannyRecord

1.2. Overview and purpose of the tool

This tool is used to record read-style speech databases. A text file with the prompts is provided to the tool. Each line of the prompt file represents a sentence or a short paragraph that has to be uttered by the speaker and stored in a separate file. The tool allows to select the sampling frequency, sample format (bits) and file format (RIFF/WAV or RAW/PCM). Several channels can be synchronously recorded. Two versions of the tool are provided: the first one uses the standard sound drivers (for windows os), and allows to record mono and stereo signals; the second version uses ASIO drivers and allows to set a multichannel configuration. For each utterance, a label file is produced including information about the prompt and recording time and date. The graphical user interface (GUI) presents the prompt text and also signal information (level for each channel, clipping). The tool also allows to use acoustic prompts: in this configuration the speaker talks after the prompt signal is played. This can be used to record mimicking a given style, or record question/answer databases. In professional recordings, an operator controls the tool: start/stop recording, repeat, advance. Graphical aspects of the tool, as font size, can be controlled to facilitate reading of the prompts from a distant position. The tool has been intensively used to produce large speech synthesis databases. For instance, the Spanish TC-STAR synthesis database, and the Catalan Festcat synthesis database, both recorded using 3 channels (2 microphones and one laryngograph), 96kHz and 24 bits/sample used NannyRecord. Also, the Catalan Speecon, a multichannel database, designed to train speech recognizers, was recorded using this tool.

1.3. A short description of the algorithm

The tool is implemented using C++. Different threads are used to attend the user/operator input that controls the tool and the speech acquisition and analysis. The input speech samples are stored in a memory buffer with a dedicated thread to store the signal in the files. The state of the buffer is monitorized to ensure that no samples are lost, even for the more demanding transfer rates (high sampling frequencies, several channels).



2. TECHNICAL INFORMATION

2.1. Software dependencies and system requirements

The tool requires windows operative system. It has been tested in Windows98, Windows XP and Windows7. The Microsoft Foundation Classes, (MFC) DLL is required. This library is usually included in windows systems. In case it is not, it can be found in Microsoft website. The multichannel version of the tool, required when more than two channels are synchronously recorded, require a multichannel audio card and ASIO drivers.

2.2. Installation

The tool does not required any installation.

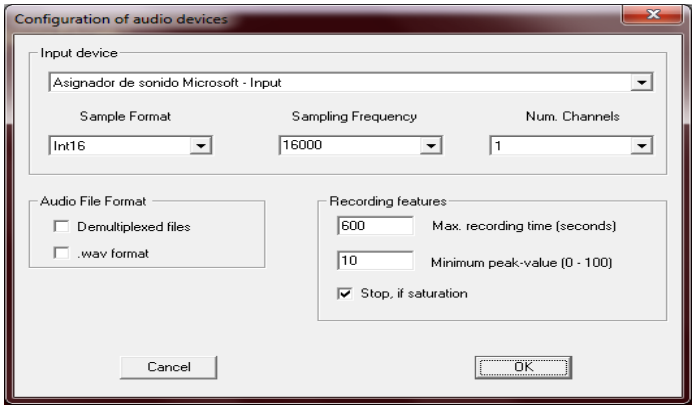
2.3. Execution instructions

The tool includes an intuitive GUI (graphical user interface). After starting the application, several buttons lauch configuration dialogs. The *configuration of audio* is used to select the recording device and the recording configuration (sampling frequency, number of channels, etc.). It is also possible to select graphical settings (background and font color, font size, vumeters), the prompt file and the output directory. During recording, several buttons and keyboard shortcuts allow to start/stop recording, navigate through the prompts (fast/forward, fast/backward), etc. All the configuration and execution settings are saved and recovered in the next session.

The audio configuracion allows to select:

- Input device
- Sample format (number of bits, etc.)
- Sampling rate
- Number of channels
- Multiplexed channels or demultiplexed channels

- RIFF/WAVE wrappers.
- Maximum recording time (after that time, the recording stops).
- Minimum peak value: if the recording is smaller it has to be re-recorder.
- Stop the recording as soon as there is clipping (saturation).



2.4. Input/Output data formats

2.5. Input data formats

The prompt file is a text file with one line for each utterance. The characters are encoded using ISO Latin-1. It is possible to indicate the filename of the output file using an optional field in the same line. In this case, the relative path and the filename (without extension) has to be included at the beginning of the line and ended with a tabulator char.

In case the acoustic prompt exist, this is played before starting the recording. The acoustic file has the same relative path and filename than the output file, but it is located in the directory "play"

2.6. Output data formats

For each utterance, a label file is generated. The extension of the file is ".txt". The text file includes the data and recording time and the prompt text.

The speech signal file can be stored in different formats: multiplexed (one file for all the channels) or different files for each channel. In multiplexed file, the samples are multiplexed in the usual way: the samples of the different channels are stores consecutively.

It is possible to select RIFF/WAV wrapper, that includes a small header with speech format information, or the headerless PCM/RAW files.

3. CONTENT INFORMATION

3.1 A test input file

Prompt file:

hamlet_dir/ham_000	<YOUR NAME>
hamlet_dir/ham_001	<DATE>
hamlet_dir/ham_002	To be, or not to be, that is the question:
hamlet_dir/ham_003	Whether 'tis Nobler in the mind to suffer
hamlet_dir/ham_004	The Slings and Arrows of outrageous Fortune,
hamlet_dir/ham_005	Or to take Arms against a Sea of troubles,
hamlet_dir/ham_006	And by opposing end them: to die, to sleep

hamlet_dir/ham_007 No more; ...

3.2. the output file

Lab file for one utterance: <DIR>\hamlet_dir\ham_002.txt
TIME: Mon Nov 26 12:56:14 2012
TEXT: To be, or not to be, that is the question:

Audio file for one utterance: <DIR>\hamlet_dir\ham_002.pcm

3.3 approximation of the time necessary to process the test input file

The tool operates in real time. Our experience with professional speakers, is that the recording sessions require approximately from 1.5 to 2.5 times the recorded duration.

4. ADMINISTRATIVE INFORMATION

4.1. Contact person

Name: Antonio Bonafonte
Address: Jordi Girona 1-3 Edifici D5, 08034 Barcelona, Spain
Affiliation: TALP research center. Universitat Politècnica de Catalunya
Position: Professor
Telephone: +34 93 401 6437
Fax: +34 93 401 6447
e-mail: antonio.bonafonte@upc.edu

4.2 Delivery medium (if relevant; description of the content of each piece of medium)

The resource will be uploaded on the MetaShare platform as an archive.

4.3 . Copyright statement and information on IPR

5. RELEVANT REFERENCES AND OTHER INFORMATION