Comparing data acquisition techniques

Lorenzo Cioni

Laboratorio di Linguistica, Scuola Normale Superiore, Pisa, Italy

The main aim of this paper is to present a qualitative and quantitative comparison among some of the techniques used at the present to acquire samples of speech. Such techniques share the source of the data as well as their final destination but they differ as to the ways by which the stream of speech is captured and stored as a stream of digital data. In the paper we will examine briefly some of the available techniques. Such techniques will be briefly examined on their own and then compared so to provide users with suggestions about the preferable technique for a certain situation.

INTRODUCTION

At the present many techniques are available for the acquisition of samples of speech as well as many are the purposes for such acquisitions but all share the source of the data (generally speaking such a source is a speaker who produces utterances that are captured by a microphone) as well as their final destination (a set of files stored on a magnetic medium) but they differ as to the ways by which the stream of speech is captured and stored as a stream of digital data.

In this paper we present the following techniques:

1. recording with a DAT and acquisition on a computer with an either analog or digital connection with an audio board driven by commercial or custom software;
2. recording with a minidisk recorder (MD) and acquisition on a computer with an either analog or digital connection with an audio board driven by commercial or custom software and
3. recording with a direct (analog) connection between a microphone and an audio board driven by commercial or custom software.

All the techniques presented in the paper make use of a Sound Blaster Live! Platinum 5.1 audio board that is equipped with both analog and digital connections [1]. It is worth mentioning that a minidisk once recorded with a MD can be loaded on a computer via an MDS-PC2 board [2, 3] but, since such a board is equipped with both digital and analog I/O connections, this case can be assimilated to the number 2 listed above.

BRIEF DESCRIPTION OF THE TECHNIQUES AND THEIR CHARACTERISTICS

In all the three cases mentioned above we have an analog phase (the microphone), an analog to digital conversion (microphone → DAT, microphone → MD or microphone → audio board) and possibly two conversions digital to analog and analog to digital if we use an analog connection between the DAT or MD and the audio board. The simplest technique is the number 3 where we have only a connection and a single analog to digital conversion. The quality of the acquisition process is determined by the quality of the microphone and of the audio board. The others are two steps techniques since they are characterized by a recording phase and an acquisition phase so that both phases determine the quality of the acquired signals.

The use of a DAT [4] allows the recording of speech in digital format usually as 16 bit linear coding with sampling frequencies of either 48kHz or 44.1 kHz with a S/N ratio better than 87 dB and a Dynamic Range better than 87 dB. The use of a MD for the recording of speech at a default sampling frequency of 4.1 kHz imposes some cautions and requires a deeper investigation since MD uses a special technique of data compression called Adaptive Transform Acoustic Coding (ATRAC).

ATRAC [2, 5] exploits some psychoacoustic principles plus sub-band coding and transform coding so to act on the frequency components of the signal and “discard” those that are perceptually irrelevant.

In this way the algorithm extracts and codes only the frequency components that are audible by the human ear. While this technique is guaranteed not to reduce sound quality and not to be perceived by listeners it is still to be proved that the frequency transformations have no influence over the parameters extracted from the speech signal.

THE EXPERIMENTAL SETTING

In order to compare the three techniques an experimental setting has been conceived. The setting is shown in the following figures 1 and 2. Figure 1 shows the first phase (recording phase). From such a figure we see how each speaker has been recorded with the three recorders in parallel so that the analysis (and the comparison) can be performed on the same utterances and under the same boundary conditions.

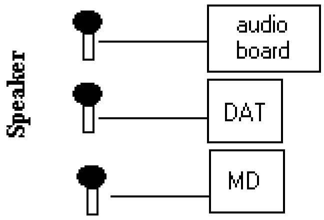

FIGURE 1. Recording Phase.

On the other hand figure 2 shows how the utterances recorded both with the DAT and the MD are stored on a Personal Computer via an audio board and either an analog or a digital (optical) connection.

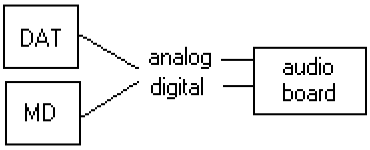

FIGURE 2. Acquisition Phases.

The acquisitions are performed with a Sound Blaster audio board driven by Goldwave using both the analog mic-in (figure 1) and line-in (figure 2) connections and the digital SPDIF-In connection (figure 2).

FREQUENCY COMPARISON

Once the signals have been acquired and stored in files they are analyzed so to compare the qualities of each acquisition chain where the microphones, the audio board and the software are the invariant elements whereas the variable elements are the recording media. The analysis of the files is mainly performed in the frequency domain since this domain is more suited to describe the transformations that the signals undergo. The frequency domain analysis include calculation of spectra (via DFT or LPC), spectrograms (via DFT or LPC), formant frequencies (via LPC) and F0. The (mainly) qualitative results over a little corpus of utterances show that the quality improves going from the microphonic recording to the DAT recording with both an analog and a digital connection to the audio board whereas it is not yet clear if the use of a MD represents a qualitative and quantitative improvement as to the aforesaid frequency parameters.

The corpus of utterances used for the frequency analysis and comparison if formed by a small set of trisyllabic and disyllabic words of the form CVCVCV, CVCCVCV (e.g. "canguro"), CVCCVCVV (e.g. "fantasia") and CVCCV /e.g. "canto") uttered in isolation.

CONCLUSIONS

In this paper we have sketched out three acquisition techniques of which we showed some characteristics. As shown in figure 1 and 2 we examined have two recordings styles and three acquisition methods. The use of a direct microphonic analog connection is well suited mainly in case of laboratory recordings with one speaker at a time. Portable DT and MD, on the other hand, can be used also in open air recordings as well as in laboratory recordings and DAT is equipped with a device that allows the use of two microphones, one on each channel, so to record two speakers at the same time. The main problem with a DAT is that speakers are aware of being recorded so that they produce at the best semi-spontaneous speech.

In such cases the use of a MD can be of help since the medium is smaller and less intrusive. The only problem is that is not yet clear if the ATRAC algorithm has a negative influence on the parameters extracted from the acquired signals.

At the end, we underline that the analysis presented in the paper is mainly qualitative and has been performed on a very small corpus produced by a single speaker. Future goals are the extension of the analysis in a more quantitative perspective and on wider corpora both acquired ad hoc or produced within other and wider research projects.

ACKNOWLEDGEMENTS

The author acknowledges the ownership of the trademarks of the products mentioned in the paper to their owners even if not explicitly cited.

REFERENCES

1. *Sound Blaster product guide,* Creative Technology, October 2000.
2. *Portable MiniDisc Recorder. Operating Instructions,* Sony Corporation, 1999.
3. *MD Editor 2. Operating Instructions,* Sony Corporation, 1999.
4. *Digital Audio Tape-corder. Operating Instructions,* Sony Corporation, 1997.
5. Tsutsui K. et alii, "ATRAC: Adaptive Transform Acoustic Coding for MiniDisc", paper presented at the 93rd Audio Engineering Society Convention in San Francisco, 1992 October 1-4.