Snorri:
An interactive tool for speech analysis

Dominique Fohr  Yves Laprie
CRIN INRIA Lorraine
BP 239 54506 Vandœuvre-lès-Nancy France

Abstract

The Snorri software has been developed at CRIN. It provides a means of editing and analysing a speech signal. The aim of Snorri is to increase the efficiency of our research in acoustic-phonetic decoding. With Snorri the user can record and play back sentences, compute and display spectrograms as well as label speech utterances. It also enables global investigation of speech labelled corpus and extraction of all the occurrences of a given phonetic pattern. Moreover, coarticulation phenomena can be studied with other tools like formant tracking displayed in the F1-F2 plane. Snorri is hence an efficient and an easy-to-use tool for speech recognition.

1 Introduction

The past ten years have witnessed a substantial development of mathematical techniques for speech decoding. Conversely knowledge based applications suffer from a dearth of reliable acoustic-phonetic information which could take into account the coarticulation effects of speech.

Thus Snorri software has been developed to provide the researcher with an efficient and easy-to-use tool for speech analysis. Two levels of use can be differentiated for Snorri. Firstly it gives facilities to record and playback a speech signal, to compute and to display spectrograms and to annotate utterances. With the second class of facilities the user can investigate a complete phonetically labelled corpus and extract the phonetic pattern he wants to study. For this purpose Snorri performs formant tracking and can display it in the F1-F2 plane.

The aim of Snorri is not to offer a set of signal processing programs like ILS [Tec 86] but to help the researcher to acquire the relevant knowledge in the speech signal. This paper describes Snorri facilities and some possible applications of this software.

2 Hardware requirements and architecture

Snorri, written in C, runs on a Masscomp 5600; it is specifically designed for efficient use of the high resolution graphic console, multi-windowing and mouse. We have upgraded the configuration with an array processor (up to 10 Mflops) which handles the signal processing (FFT, LPC, cepstrum). All signal processing algorithms are nevertheless available without recourse to the array processor. Snorri also has the capability of generating hardcopies for Postscript printers. We are currently working to implement Snorri on Sun stations with X multi-windowing.

The user can recompose the screen layout by calling up the multi-windowing capabilities. He can access all the facilities via menus. The menus contain the functions to manipulate a representation of speech (temporal signal, spectrogram) or the functions to accomplish a specific task (e.g. phonetic annotation).

In order to access all the available speech signal files, the user can describe the speech signal file format he wants to use. Snorri can then read all the speech files according to the specified format which contains the following information: number of bits per sample, sampling rate, sample swap indicator, offset of the beginning of speech relative to file beginning, size of speech signal... Commands of format file can give values for these parameters or the way to read them directly in the speech signal file. No limits are imposed for the size of speech signal files and Snorri can handle files with several sentences (as in BDSONS database).

3 Snorri tools

3.1 classical tools

Recording  The user can choose the sampling rate to record a sentence (the default sampling rate is 16 kHz). Snorri automatically finds the beginning and the end of the utterance; moreover it is possible to move boundaries to pay attention to a part of the speech signal. Snorri enables the sentence to be played back and provides a zoom to examine the speech signal as accurately as possible. The mouse can be used to choose a file in the database when using a corpus.
Spectrograms
Snorri computes and displays high quality spectrogram with a 13 colour map which the user can redefine. The standard spectrogram is a wide-band spectrogram (4 ms Hamming window with a 2 ms shift) but Snorri can also generate the following spectrograms:

- spectrograms computed on LPC coefficients which enhance formant frequencies,
- narrow-band spectrograms (32 ms Hamming window) in order to separate the fundamental harmonic frequencies,
- cepstrally smoothed spectrograms (with cepstral filtering of \[Rah 69\]); the user can adapt smoothing according to the speaker's fundamental frequency.

The algorithms make use of the array processor; so the time of computation + display is short (0.5 second for computation of a 2 second spectrogram and 8 seconds for displaying).

We have chosen to set the default duration of spectrograms on 2 seconds so that the user does not have to adapt to several display formats. The user can move the 2 ms Snorri window anywhere in a speech signal file whose duration exceeds 2 seconds. As spectrogram display and temporal signal are synchronized the user can escape the 2 ms window simply by zooming on a part of the temporal signal and then displaying the associated spectrogram (Fig. 1).

Figure 1: high resolution spectrogram

Phonetic annotation
One of the first tasks for speech scientists is to label sentences of the available corpus. Thus we have given Snorri all the capabilities for this work. The phonetician can put segment boundaries either on the main spectrogram or on the zoom; he can listen to the segment which he is labelling. He has at his disposal tools to modify or to cancel any segment he has positioned on the speech signal. Other facilities enable movement anywhere in the utterance and to find locations of existing labels.

3.2 More sophisticated tools

3.2.1 Semi-automatic phonetic annotation

The knowledge based acoustic-phonetic decoder Aphodex, implemented by D. Fohr [Car 87] takes advantage of the Snorri facilities. This also allows the efficiency of speech labelling to be increased considerably. It is in fact a tedious task for a phonetician expert to segment and label speech sentences. In addition, it tends to introduce human error. The coarse segmentation performed by Aphodex enables time alignment using time warping techniques to be carried out.

The user enters the phonetic transcription of a given sentence which is segmented by Snorri and then aligned according to the transcription. During alignment, only the phonetic classes (plosives, fricatives, vowels and sonorants) are taken into account. Phonetic annotation is then much faster since the user only has to correct imprecise boundaries without choosing phonetic labels. Moreover errors stemming from erroneous phonetic label choices disappear.

3.2.2 Investigation of labelled corpus

The major aim of Snorri is to increase our knowledge of speech, and to develop more reliable methods of extracting relevant acoustic features. It is with this aim in mind that we have designed accesses to labelled corpora. In order to distinguish between variations associated with the speaker and these associated to context a speech analysis system has to enable fast accesses to all the occurrences of a given phoneme belonging to a corpus as well as its contextual phonemes.

Sentences can be stored either in corpora which are common to all the speech researchers working with Snorri or in corpora which are directly devoted to a given study. This has led us to choose the following organisation:

- The user has the choice between the public corpus and his own one. A simple mouse click allows the user to choose the speech signal file he wants to work on.
• Phonetician can extract either a given phoneme (e.g. /i/) or a phoneme sequence (e.g. /abi/) or a composite phoneme or phonetic class sequence (e.g. front vowels preceded by /b/) in the whole labelled corpus. After the user has defined one or several phoneme sequences Snorri scans the corpus to extract all the occurrences belonging to the corpus with the associated labelling. Snorri build a speech signal file and the corresponding label file. Both files are added to the private corpus and facilitate an efficient way of discovering reliable acoustic correlates by displaying the same phoneme sequence for a large set of speakers.

Fig. 2 shows the extraction of the sequence /labial/ + /i/.
We can conclude from this figure that two obvious cues describe the sequence /labial/ + /i/:

• F2 and F3 rise sharply at the beginning of the vowel,
• F2 appears faster than F3.

This facility also offers a very convenient way of testing and debugging our speech recognition algorithms in a given limited phonetic context.

3.2.3 Formant tracking
Most phonetic knowledge makes use of formant trajectories. On the one hand formant frequencies describe the target phonemes and on the other hand formant trajectories are a good means of taking into account contextual effects.

Thus Snorri provides a formant tracking algorithm. After preprocessing, our algorithm (outlined in [Foh 88]) can work either on LPC roots or on cepstrally smoothed spectra, and uses the events (fusion, split, loop of peak lines, omission of one peak) it encounters.

The phonetician often needs to characterize coarticulation phenomena in terms of proximity to the expected target phonemes. Thus Snorri displays formant tracking either on the spectrogram or in the F1-F2 plane where average vowels for a male French speaker are shown. Fig. 3 shows the formant tracking path when the speaker pronounces /i//a//u/.

3.2.4 Other tools
Snorri provides the user with other tools:

• energy in any frequency band,
• narrow-band spectrum slice, smoothed narrow-band spectrum slice, LPC and cepstrally smoothed spectrum slice (Fig. 4) for a signal window the user can designate with the mouse,
• pitch detector with Shafer and Rabiner algorithm [Rab 78].

4 Snorri and the other speech analysis systems
Snorri is a specialized tool for speech study; it provides the basic set of capabilities for speech scientists and enables phonetic knowledge acquisition without programming. In
this sense, Snorri is quite different from ILS, which is a collection of signal processing routines.

Nevertheless using powerful graphic tools and computation hardware, Snorri enables investigating of a large amount of speech.

Figure 4: From the bottom up: narrow-band, smoothed narrow-band, LPC and cepstrally smoothed spectra

Among all the details which differentiate Snorri from other systems we consider that the most important point concerns its ability to exhibit contextual influences:

- extraction of a phonetic pattern from annotated corpus.
- formant tracking is displayed in the F1-F2 plane.

Although Snorri has not been designed as an open system, the user can readily implement his own application using the basic facilities provided by Snorri. Since Snorri offers most tools for speech recognition (formant tracking, F0, energy ...) D. Fohr has developed the expert system in spectrogram reading Aphodex [Car 87]. B. Mangeol has added a lexical component using Aphodex results [Rom 88].

5 Concluding remarks

Snorri has been designed with the conviction that continuous speech recognition can be achieved with a vast knowledge of speech phenomena. This also implies undertaking a considerable work of learning and requires tools that enable this work. Snorri is a first leap forward:

- It is an easy-to-use software which does not require any programming experience.
- It allows a researcher to access a very large amount of speech data and to focus on a given acoustico-phonetic issue. Furthermore it can generate data to test a specific aspect of speech decoding.
- Snorri has been designed with the advice of phoneticians who have a large experience of speech labelling and other speech analysis systems.

In order to continue in this direction we will add to Snorri tools to express and extract acoustic-features in the speech signal. This can be carried out either by giving the researcher statistical analysis tools to study a given acoustic parameter (as in [Shi 83]) or by giving Snorri full automatic learning capabilities whose results are judged by a phonetician.

It is also important to share and develop available tools in speech study with other researchers who may find Snorri helpful and can extend and improve its capabilities. We have thus given Snorri to other speech groups mainly in France (Cerfa, CNET, Matsushita ...).

References


