

Snorri, a Software for Speech Sciences

Yves Laprie

LORIA - UMR 7503 Speech group
BP 239 54506 Vandœuvre-lès-Nancy FRANCE
Yves.Laprie@loria.fr

Abstract

Tools for investigating speech signals are an invaluable help to teach phonetics and more generally speech sciences. For several years we have undertaken the development of the software WinSnorri which is a research tool and an illustration tool for both speech scientists and teachers in phonetics. It consists of five types of tools:

- to edit speech signals,
- to annotate phonetically or orthographically speech signals. WinSnorri offers tools to explore annotated corpora automatically,
- to analyse speech with several spectral analyses and monitor spectral peaks along time,
- to study prosody. Besides pitch calculation it is possible to synthesise new signals by modifying the F0 curve and/or the speech rate,
- to generate parameters for the Klatt synthesiser. A user friendly graphic interface and copy synthesis tools allows the user to generate files for the Klatt synthesiser easily.

In the context of speech sciences Snorri can therefore be exploited for many purposes, among them, illustrating speech phenomena and investigating acoustic cues of speech sounds and prosody.

1. Introduction

WinSnorri is a software aimed at assisting researchers and teachers in the field of speech sciences. WinSnorri was designed with two general goals in mind. First, it provides a set of capabilities such that speech scientists as well as students with little or no programming experience are able to analyse and edit speech signals. Second, it provides a framework for teachers to illustrate speech phenomena and show the perceptive importance of acoustic cues.

The tools provided by WinSnorri were therefore designed with a principal concern of making their use on speech signals very easy. This explains and justifies the overall organisation of the software and the set of default parameters retained, such as those needed for the calculation of spectrograms.

Snorri was originally developed on a Masscomp machine in 1987 [1]. Then, it was adapted to X11 with the Motif toolkit for Unix machines before being ported on Windows. Snorri is the name of the Unix version and

WinSnorri that of the Windows (95, NT, 98) version. Snorri was distributed with sources to most of the French laboratories working on speech.

2. What does WinSnorri ?

WinSnorri allows users to edit speech signals, calculate and display spectrograms, annotate phonetically or orthographically speech signals, compute dynamically the results of various spectral analyses, explore automatically speech corpora, monitor formant trajectories, pilot the Klatt synthesiser, extract and modify F0 and speech rate.

We pay a particular attention on data journaling, i.e. the possibility of extracting results within WinSnorri and put them in an ASCII text file. All the results (except the spectrogram and speech signal which would occupy too much space and present little interest as is) are automatically journaled as ASCII output by WinSnorri. The user can append these results on a log file for subsequent processing with other software.

2.1 Speech editor

Editing facilities are the basis of WinSnorri. They have been designed to help speech scientists in creating speech stimuli rather than manipulating very long signals, even if there is no limit on the length of the speech signals. The spectrogram is always displayed because it is undoubtedly the best tool to evaluate acoustic consequences of editing commands. This, in particular, allows the user to discover spectral "clicks" stemming from the suppression of the quasi-periodic structure of speech. Besides basic facilities (cut, paste, copy...) we added filtering and attenuation commands. Both of them can be used directly on the spectrogram by pointing the time x frequency region the command is run on (WinSnorri automatically designs the filter according to the frequency region to be filtered and the attenuation level).

In order to keep the same visual time resolution for spectrograms the default length of WinSnorri window is set to 4 seconds. WinSnorri displays large band, narrow band, linear prediction and cepstrally smoothed spectrograms. The length of analysis windows as well as the range of the energy scale can be changed by the user. A draft mode allows the user to inspect a speech signal very rapidly in order to localise a characteristic acoustic event.

2.2 Spectral analyses

The user can open a window to display spectral slices such as Fourier transforms, simple or selective linear prediction and cepstrally smoothed spectra. It is possible to adjust all the spectral parameters. Spectra are themselves calculated by following the movement of the mouse on the spectrogram. The user can also open a text file into which numerical results are saved. In addition, WinSnorri can display the peaks of cepstrally smoothed spectra or the roots of linear prediction polynomials on the spectrogram and thereby highlight formants.

2.3 Phonetic and orthographic annotations

WinSnorri provides the user with the possibility of both phonetic and orthographic annotating. With the mouse and a menu containing the list of the phonemes of the language under investigation the user constructs phonetic annotations for the speech signal. The list of phonemes can be modified at will, thus making WinSnorri multilingual with the possibility of application to any set of phonemes or to any type of marks irrespective of the language considered. Furthermore, WinSnorri offers complementary editing and file management tools such as displacing, deleting and searching, finding phonetic and orthographic labels, as well as reading and saving annotation files.

The set of phonemes can be customised by the user which allows any phonetic or non phonetic font to be used. Besides phonetic annotations it is thus possible to specify a set of non phonetic symbols, prosodic symbols for instance, to annotate speech signals.

There are two ways of annotating speech signals phonetically. The first is a hierarchical phonetic menu which helps the user in finding the right phonetic symbol. Once the user has the phonetic symbol set in mind he can use the second mode of annotating which simply consists of a dialogue window containing the set of phonemes. The user can enter annotations with the keyboard or by clicking on phonetic symbols. This is substantially faster than dragging the mouse through the hierarchical phonetic menu. This facility has required a narrow collaboration with phoneticians to reach a sufficient level of ergonomics.

Numerous speech corpora are now available. Some of them are phonetically annotated and represent therefore an invaluable source for investigating specific acoustic cues. WinSnorri can read most of the annotation formats which allows most of the corpora to be explored. After the user has specified several sequences made up of phonemes and/or phonetic classes, WinSnorri builds a speech file and an annotation file with all of the occurrences of these sequences found in the corpus under investigation. This exploration facility is also available for orthographically annotated corpora.

2.4 Prosody

Tools to studying prosody rely on a pitch tracking algorithm based on a spectral comb method [2]. Voicing decisions as well as F0 value corrections are achieved

by a dynamic programming algorithm which selects F0 values to be accepted as results. Besides F0 calculation, WinSnorri includes synthesis capabilities to modify prosody. We used the TD-PSOLA (Time Domain Pitch Synchronous Overlap and Add) method [3] which has the advantage of being very fast and easy to implement. Nevertheless this method requires the decomposition of the original signal into overlapping windows synchronised with pitch periods. We therefore implemented an algorithm which extracts pitch marks from the speech signal. Windows used by TD-PSOLA are centred on these marks. This algorithm [4] uses dynamic programming to find out a set of signal extrema spaced by pitch periods. Speech rate modifications as well as F0 modifications amount to space out windows according to the new values of F0 or, to duplicate or eliminate windows according to the new speech rate.

There are two kinds of modification. First, the user can change the speech rate and/or the F0 level for the whole speech signal. Second, the user can change the F0 contour by editing the F0 curve which is superimposed on the spectrogram. A new signal is then re-synthesised according to the modifications.

Although WinSnorri includes the calculation of the energy in any frequency band it is not possible yet to edit the energy contour and then synthesise a new signal accordingly.

2.5 Klatt synthesiser

Having at disposal a synthesis system which easily controls acoustic cues is an important advantage for speech scientists who are concerned with speech perception. Snorri (and in the near future WinSnorri) includes a graphical interface for the Klatt 1980 formant synthesiser [5]. Most of the work in constructing stimuli consists of choosing the 39 parameters from existing speech signals. We therefore pay a particular attention to develop tools to realise copy synthesis. To our knowledge, although some copy-synthesis systems have been developed for other formant synthesisers, there does not exist such a system for Klatt's synthesiser.

WinSnorri includes tools to draw parameters on the spectrogram. As it is really difficult to track formants by hand on the spectrogram the user can draw rough formant trajectories and then "register" them onto the spectrogram peaks (or on the roots of linear prediction). Other editing facilities to merge, destroy, move or smooth trajectories are available.

The main issue in copy synthesis is how to set formant bandwidths in the case of the cascade branch, bandwidths and amplitudes in the case of the parallel branch of the synthesiser. Preliminary experiments brought to light that is preferable to use only the parallel branch of the synthesiser because this allows both vowel and non-vowel sounds to be synthesised. One of the advantages of the parallel branch is that formant amplitude can be specified independently of the bandwidth. We therefore accepted the parallel branch to

perform copy synthesis. As the bandwidth is difficult to measure we set it to a constant defined according to the formant (from 50 Hz for F1 to 140 Hz for F3). The amplitude of formants, therefore, depends only on the amplitude parameter.

The copy synthesis function determines the amplitude of formants so that synthetic speech approximates the original signal at best at each instant. Amplitudes are measured from the spectrum of original speech and that of the excitation signal. The quality of copy-synthesis depends on the ability of the spectral analysis to fit the peaks of the original spectrum. We accepted a method derived from cepstral analysis, called "true envelope" [6], which iteratively corrects cepstral coefficients so that the contribution of spectral values above the smoothed spectrum (mainly harmonics) is reduced. This method yields very good results for original speech and excitation spectrum as well.

The standard scenario to copy an original utterance is as follows. The user draws formants trajectories, register them, smooth them if needed, recover the F0 contour and computes formant amplitudes by means of the copy synthesis procedure. The signal analysis strategy used to set amplitude begins by determining excitation parameters. Once the fundamental frequency has been calculated, the frication amplitude is set to 60 dB for unvoiced sounds. For voiced sounds the mixture between voiced and noise source is set according to the ratio of energy in low and high frequency regions, so that higher formants correspond to noise, especially in voiced fricatives and bursts.

The graphical interface of the synthesiser can also be used in a simpler way to monitor formant trajectories and save the results in the form of a file including formant frequencies and amplitudes among the 39 parameters of Klatt. Formants F1 and F2 can be displayed in the F1-F2 plane in order to evaluate the proximity of their trajectories with vowels of the language under investigation.

3. How phonetic teachers can use WinSnorri ?

To our knowledge WinSnorri has been used in four ways:

- to illustrate speech phenomena,
- to study prosody,
- to study coarticulation,
- to prepare stimuli for perception experiments.

3.1 Illustration of speech phenomena and speech analysis

Undoubtedly the first interest of WinSnorri lies in the facilities to "see" and analyse speech signals. This allows teachers to prepare courses in spectrogram reading, possibly in the form of hypertext documents including utterances displayed with WinSnorri.

In addition to illustrating speech, WinSnorri allows speech signals to be analysed and results extracted from WinSnorri by means of data journaling to be exploited with other software.

3.2 Study of prosody

WinSnorri includes a complete toolbox to studying prosody which can be used to illustrate prosody as well as to show the relation between perceptual effects and F0 variations (see [7] for more details).

3.3 Study of coarticulation

WinSnorri provides the user with numerous tools ranging from displaying LPC roots or spectral peaks to the graphical interface of the Klatt synthesiser. Students are thus given the possibility of evaluating the extent of the coarticulation phenomena and their variability with respect to the speaker by exploring phonetically annotated corpora.

3.4 Preparation of stimuli for perception experiments

This is probably another strength of WinSnorri which gives the user numerous tools to edit speech signals and create artificial stimuli, the timbre of which is close to that of natural speech. There are three kinds of tools: editing tools (basic facilities completed by filtering and attenuation), graphical interface of the Klatt synthesiser with the copy synthesis function, and prosody modification.

These tools can be exploited by students to evaluate the contribution of certain acoustic cues to speech perception, they can also be used in a more advanced way to study speech-speaker dichotomy with the Klatt synthesiser or the perception of bursts of stops consonants.

4. Perspectives

The goal of WinSnorri is the same as in the Spire projects [9], i.e. to create a research environment that is easy to use.

It is planned to continue to add new facilities in WinSnorri, among them a phoneme aligner based on automatic recognition systems developed by the Speech group at LORIA.

We are convinced that speech analysis can give rise to efficient tools to teaching foreign language phonetics. This issue should be investigated with the help of teachers in phonetics and foreign language. We will therefore add a language of macro functions intended to create scenarios that could be used as exercises.

Acknowledgements

We would like to thank C. Renard, J. Vaissière, D. Fohr, C. Antoine, A. Bonneau and members of the speech group at LORIA who contributed to various aspects of the development of Snorri and WinSnorri.

References

- [1] D. Fohr and Y. Laprie (1989). Snorri: an Interactive tool for Speech Analysis, *Proc Eurospeech'89*, Paris, Sep. 1989.
- [2] P. Martin. (1982) Comparison of Pitch Detection by Cepstrum and Spectral Comb Analysis, *Proc. of Int. Conf. Acoust. Speech and Signal Processing*, 1982, 180-183.
- [3] E. Moulines and J. Laroche (1994) Non-parametric techniques for pitch-scale and time-scale modification of speech, *Speech Communication*, 16, 175-205.
- [4] Y. Laprie and V. Colotte (1998). Automatic pitch marking for speech transformations via TD-PSOLA, *Proc of the European Signal Processing Conference*, 1133-1136, Rhodes, Sep. 1998.
- [5] D.H. Klatt (1980). Software for a cascade/parallel formant synthesizer, *J. Acoust. Soc. Am.*, 67, 971-995.
- [6] P. Halle (1983) Techniques cepstrales améliorées pour l'extraction d'enveloppe spectrale et la détection du pitch, *Actes du séminaire « Traitement du signal de parole »*, 83-93, Paris, 1983.
- [7] A. Bonneau, Y. Laprie and J. Vaissière (1999) Hypertext atlas of speech, *Proc. Matisse Workshop on Method and Tool Innovations for Speech Science Education.*, 1999, London.
- [8] A. Bonneau, S. Coste, L. Djezzar and Y. Laprie, (1992). Two Level Acoustic Cues for Consistent Stop Identification, *Proceedings International Conference on Spoken Language Processing*, 511-514, Banff, 1992
- [9] V. W. Zue, D. S. Cyphers, R. H. Kassel, D. H. Kaufman, H. C. Leung, M. Randolph, S. Seneff, J. E. Unverferth and T. Wilson (1986). The development of the MIT Lisp-Machine, *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, 329-332, Tokyo, 1986

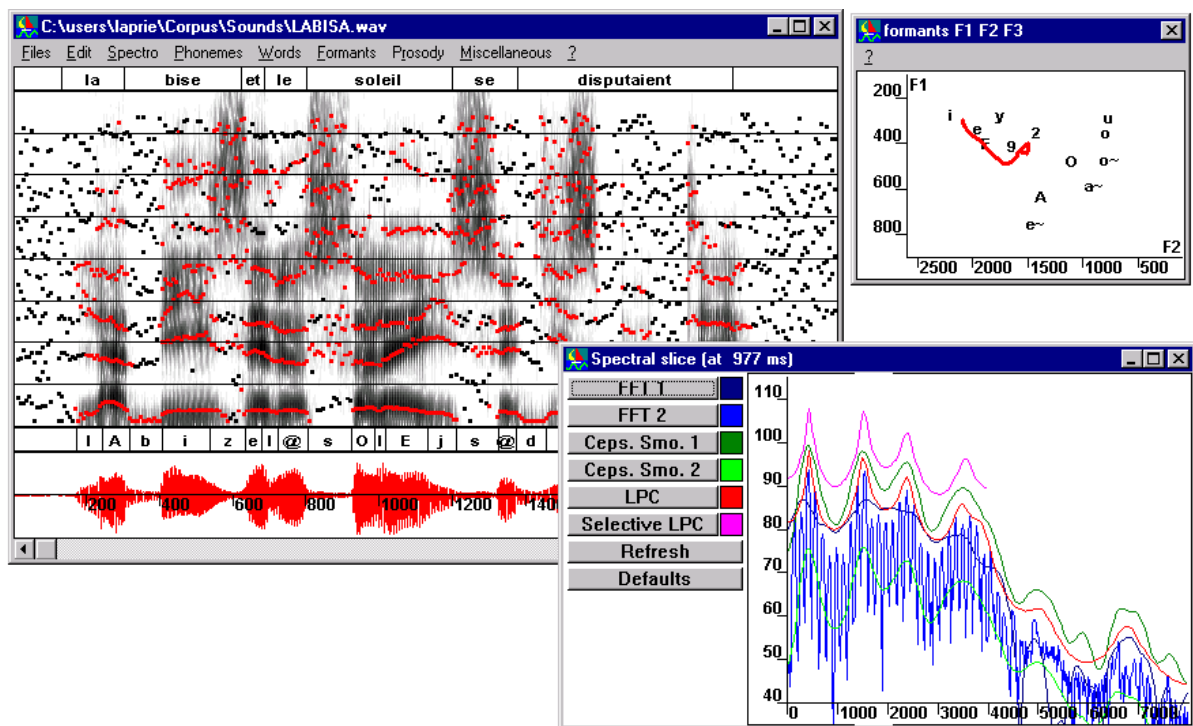


Fig. 1 : General layout of WinSnorri (spectrogram, F1-F2 plane and spectral slice)