WinPitch, a multimodal tool for speech analysis of endangered languages

Philippe Martin

EA3967 CLILLAC-ARP, UFR Linguistique
Université Paris Diderot Paris 7, Paris, France
philippe.martin@linguist.jussieu.fr

Abstract

WinPitch is a speech analysis program running on PC and Mac personal computers for acoustical analysis of speech corpora. It includes a large number of specialized functions to transcribe, align and analyze large sound and video recordings. It supports multiple hierarchical layers for segmentation (up to 96 layers), speaker lists, and overlapping speech. Various character encodings, including Unicode, are supported, with optional right to left text display for Arabic and Hebrew transcriptions. Interfaces with other popular speech analysis programs are provided, as well as standard alignment input and output in XML format. Many functions are devoted to the transcription, alignment and description of less documented languages, such as slow speed playback, programmable keyboard, automatic lexicon generation and text labeling. Various software functions are described together with their applications to the analysis of Parkatêjê, a Timbira language spoken in the Amazonia by about 400 speakers.

Index Terms: speech analysis, transcription, alignment, Parkatêjê, endangered languages.

1. Introduction

WinPitch is a software program devoted to acoustic analysis of speech which includes, as its name suggests, specialized functions for research in prosody. It has been continuously developed since 1995 and runs under Windows (any flavor) on PC and Mac personal computers. Many original functions allow effective acoustical analysis of large scale speech corpora, as demonstrated in its use in the C-ORAL-ROM project [1], which assembled transcribed and aligned large spontaneous speech recordings dealing with similar topics in French, Italian, Spanish and European Portuguese. In a project directed by L. Araujo [4], pertaining to prosodic analysis of Parkatêjê, a Timbira language spoken in Amazonia, many dedicated functions were integrated into the software to allow for an easy acoustical analysis of prosodic features of this language, and more particularly the syllable stress distribution.

The program screen is divided in 4 sections: 1) a command section, with specialized windows grouping the essential parameters related to a particular function (e.g. recording, playback, prosodic morphing, transcription, alignment, statistical analysis, etc.), 2) a navigation window (displaying the wave form), 3) an analysis windows, displaying a spectrogram, fundamental frequency and intensity curves as well as the waveform related to the speech section selected in the navigation window, and 4) an data retrieving window displayed text in aligned segments for easy retrieval of corresponding speech data (Fig. 1).

WinPitch can handle stereo signals and display the resulting analyzed parameters in different colors. The program can also analyze multimedia recordings (many video formats are supported, such as avi, mp4, wmv, flv, etc.) while keeping functional all the other features, such as reduced speed speech playback.

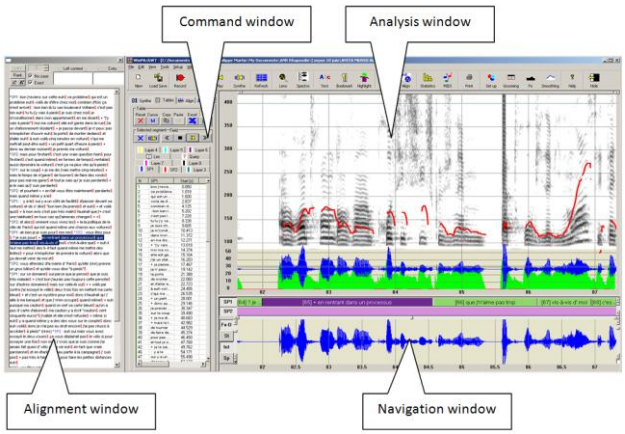

Figure 1: WinPitch command, alignment, navigation and analysis windows

2. Sound recording made clear

Among the unique features not found in other popular programs such as Transcriber [2] or Praat [3], we can mention real-time spectrographic display, allowing visual monitoring of speech recordings. This is especially useful as research speech corpora are rarely recorded by sound engineers, which often results in poor sound quality recordings (background noise, echo, wrongly adjusted recording level, microphone filtering, etc.). Poorly recorded speech samples can make syllabic prominence and fundamental frequency analysis difficult or impossible. Since most of personal computers contain a sound card, it is very easy to implement a speech monitoring system by merely adding an appropriate microphone while running WinPitch. This function was especially useful when recording in the Amazonian forest, as background noise (not due here to neighbor passing cars or room echo, but to close by singing birds or other animals)

3. Programmable keyboard

An orthographic system has been designed in 1997 by L. Araujo [4], using a mixture of symbols taken from the Latin 1

supplement, Latin Extended A, Latin Extended B font families. These symbols being available together in a Unicode font, it is possible with WinPitch to specify a dedicated keyboard using all the necessary symbols to easily transcribe a Parkatêjê recording with the standardized orthographic system.

Phonème	Orthographe	Exemple
p	p	*Pàr* "arbre" *tep* "Poisson"
t	t	*Tîr* "vivant" *pɣt* "soleil"
č	x	*Xô* "fruit" *kaxêr* "lune"
k	k	*Kô* "eau" *arîk* "se taire"
ʔ	h	*Âhãre* "poule" *ihpa* "son bras"
m	m	*Mîre* "crocodile" *mɾa* "pleurer"
n	n	*nã* "donc" *pàn* "ara"
ɾ	r	*Rop* "jaguar" *hahêr* "mur" *krô* "sanglier"
w	w	*Wai* "xamã" *kwỳr* "manioc"
y	j/i	Pjê "boue" *paipai* "vomir"
h	h	*Hõr* "dormir" *põhy* "maïs"
a	a	*Ipa* "mon bras" *kahã* "couleuvre"
ə (nasal)	ã	*Irã* "fleur" *prãm* "avoir envie"
ɛ	e	*Tep* "Poisson" *ite* "jambe"
e	ê	*Mpoxê* "fil" *inxê* "mère de EGO"
ë	ë	*krë/ krër* "manger"
i	i	*Krir* "fourmi"
ĭ	ĭ	*Krĭ* "village"
ɔ	o	*Popok* "canne fine"
ő	ő	*Atõ* "frère"
o	ô	*Kô* "eau" *tôn* "tatou"
u	u	*Pur* "abati"
ü	ü	*Küm* "fumée"
ï	y	*Mpohy* "pépin"
Ĭ (nasal)	ỹ	*hỹ* "oui"
ë	ỳ	*kwỳr* "manioc"
a	à	*Jàt* "patate douce"

Table 1 : Orthographique conventions (from Araujo [4])

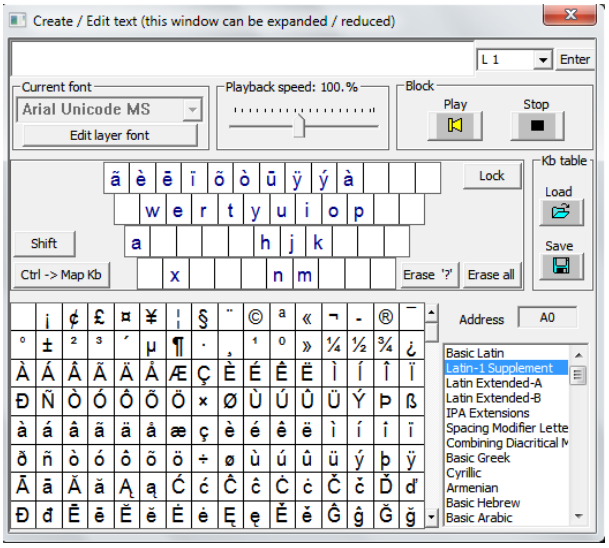

Figure 1. User designed specialized keyboard for Parkatêjê orthographic transcription. Once defined, the special purpose keyboard layout can be saved and retrieved for the next work session.

4. Transcription and alignment

Aside from classical transcription tools (with automatic segmentation in short sections, automatic segmentation in syllables and phones and user defined variable playback speed), WinPitch has a unique function allowing easy alignment of recordings already transcribed but not aligned, as frequently found in on line corpora or elsewhere.

This function is especially useful in case of poorly recording examples, where automatic alignment is ineffective. It allows the user to click on any unit of text (whether on words, syntagm or whole sentences) while the speech is played back at user selectable reduced speed (down to 7 times real-time, Fig. 2).

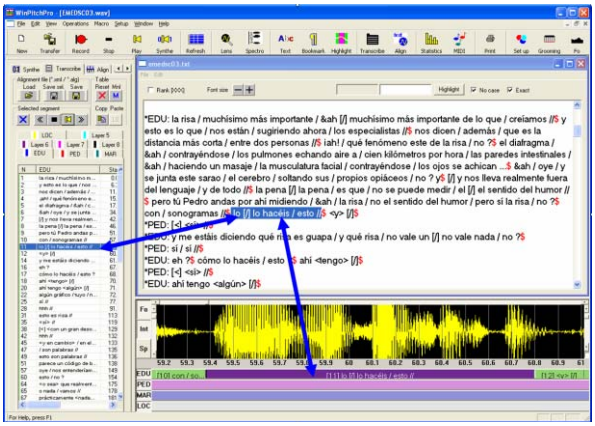

Figure 2. Assisted alignment by slowing down speech playback. At each mouse click on a unit of text perceived at slower speed (top right window), bidirectional pointers are generated automatically between the corresponding speech segment (bottom right window) and the text database (left window).

This process allows an easy and close to real time alignment of already transcribed text even for poorly recorded examples, since the difficult task of automatic speech recognition is passed to the more efficient human recognition system. The whole method precludes a time consuming segment by segment alignment if the speech transcription is available but not aligned. Besides, other WinPitch modes of transcription include automatic segmentation based on silence or pause boundaries, where the user enters directly the corresponding text of predefined segments.

Manual fine tuning of on the fly alignment can be easily accomplished by displaying an underlying narrow band spectrogram, which also can be used to align overlapping sections of speech (Fig. 3).

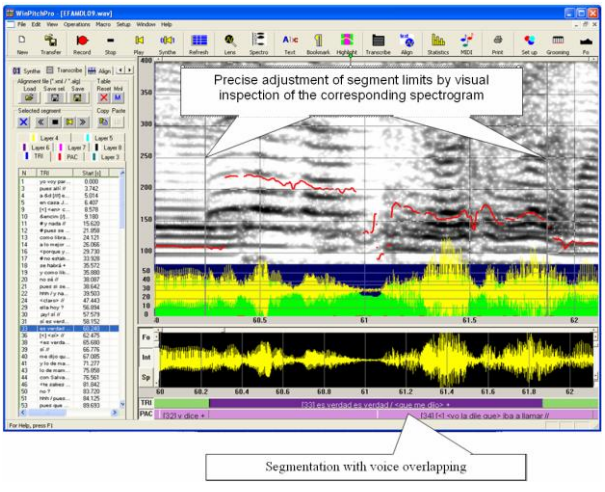

Figure 3. Fine tuning of speech segments limits with the help of a simultaneously displayed spectrogram (which allows precise segmentation in case of speaker's overlapping).

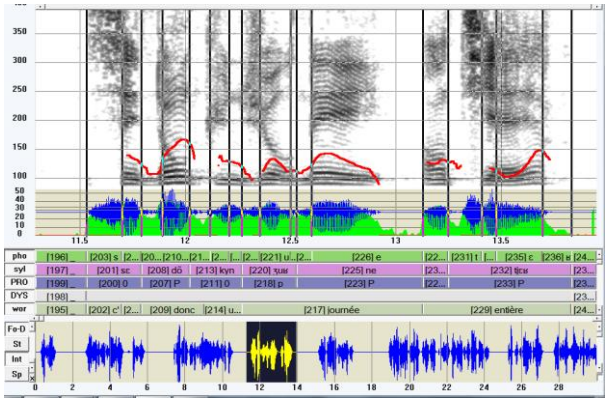

Figure 4. Automatic API transcription from orthographic text and morphological and syntactic labeling.

5. Data mining

Transcribed and aligned data can be easily retrieved

1) by merely selecting with the mouse the desired section of text;
2) by selecting an entry in a dynamically build lexicon containing all text entries (including API transcriptions or morphological and syntactic labeling generated automatically);
3) by using a table of text segments, or 4) by entering the researched text with its optional left and right context. Low pass and high pass filtering is also provided to implement perceptive (or other) tests.

Lexical items can be searched with or without their left and right contexts, the sequence of found occurrences extending automatically to a practically unlimited set of transcription files placed in a common directory.

Native output format is in XML (for alignment files) and proprietary WP2 format (which includes all the annotations, text, highlighting, Fo tracking parameters, etc. as defined by the user).

WinPitch includes also an automatic prominence analyzer operating from built in automatic syllabic detection or from a manual or imported syllabic segmentation. An automated consulting tool is integrated in the program for automatic syntactic and morphological labeling as well as an API transcription from data extracted from large lexicon in Excel format (Fig. 6). Successive occurrences taken form an unlimited set of transcribed files of a user specified item are obtained with a single mouse click, together with their sound acoustic analysis and syntactic context.

6. Acoustic analysis

Since pitch tracking algorithms are so far prone to errors in adverse recording conditions, and given that for a particular speech segment some algorithms are less prone to errors than others, WinPitch includes 6 different pitch tracking routines to evaluate the fundamental frequency (spectral comb, spectral brush, autocorrelation, AMDF, spectral fit, harmonic selection).

These algorithms and their related parameters can be independently applied on user defined segments of the speech wave, in order to use the most appropriate scheme in a given speech section of the recording. The spectral comb and spectral brush are especially resistant to noise and absence of some harmonics in the spectrum, so that even creaky segments can be adequately analyzed with appropriate parameters (Fig. 7). WinPitch includes also a scanning feature allowing a quality analysis of the recording in terms of fundamental frequency coherence, transition and presence of creak, so that the user can easily retrieve speech segments with Fo tracking problems.

7. Automatic labeling

Another interesting feature of the software, devoted more specifically to prosodic research, is the prosodic morphing tool, where fundamental frequency, intensity, and syllabic duration can be easily modified with simple and intuitive graphic commands (Fig.8). The syllabic (or phone) durations for example can be modified by a single mouse move after automatic or manually defined syllabic or phone boundaries (imported for example from a Praat TextGrid file).

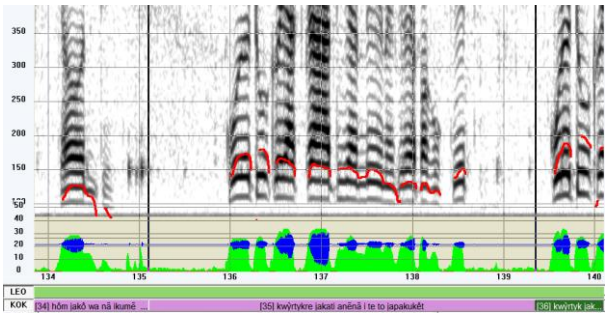

Figure 5. Prosodic morphing with modifications of fundamental frequency and syllabic duration.

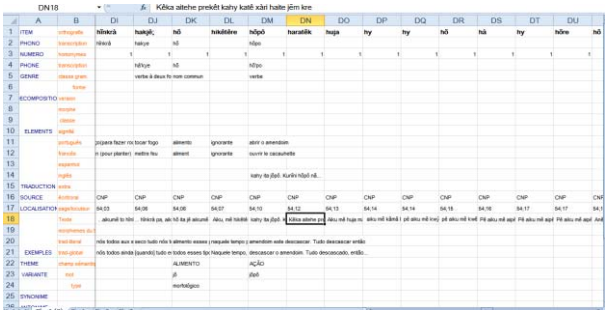

Figure 6. Lexicon in Excel format giving for each Parkatêjê entry its phonetic and phonologic transcription, its morphological composition, its lexical category and variants as well as the translations in French, English, Spanish and Portuguese.

8. WinPitch interfaces with other programs

WinPitch can import Transcriber, PFC, Necte files among others, and read and save Praat files (old and new TextGrid format). All data can be exported in Ascii (with Unicode extension) directly as a text file or into Excel® with one mouse click. It can load wav and many other sound or video files directly, with resampling into any user selected sampling frequency. This is especially important to avoid wasting storage space and computing power by using a too high sampling frequency, whereas 16,000 Hz or 22,050 Hz are sufficient for speech recordings.

Sound files can be edited (segment deletion, copy and paste), and can be concatenated or "glued" together to form a stereo file from 2 mono files (in case where a same event recorded into two independent files must be analyzed together). Text can be added (in any color and font) on the analysis window for illustration purposes to be included in a research paper. The resulting augmented analysis window can then be exported in a picture format in a text editor such as Word for example. Segments of the acoustic analysis (Fo, intensity, waveform, spectrogram) can be highlighted and independently labeled, for paper illustration and for later selection in Excel (or other program) for further statistical analysis.

9. WinPitch as shareware

The software program is presently a shareware, free for the asking and downloadable from www.winpitch.com. As, contrary to other popular programs, it never received any public funding, a small contribution is expected from interested institutions.

References

[1] C-ORAL-ROM (2005) Integrated Reference Corpora for Spoken Romance Languages, Edited by Emanuela **Cresti** and Massimo **Moneglia,** Studies in Corpus Linguistics 15, John Benjamins, Amsterdam. (http://lablita.dit.unifi.it/coralrom/).

[2] Transcriber, a tool for segmenting, labeling and transcribing speech, http://trans.sourceforge.net/en/presentation.php

[3] Praat, www.praat.org.

[4] Araújo, L., 1993, "Fonologia e grafia da língua da comunidade parkatêjê (timbira)" in: Luci SEKI (org) *Lingüística indígena e educação na América Latina*, Campinas/SP, Editora da UNICAMP, pp 265-271.

[5] WinPitch, 1996, 2011, http://www.winpitch.com