

Carla Simões,  
t-carlas@microsoft.com

**Microsoft** | Development Center  
*Portugal*



# Speech Analysis and Transcription Software

- Methods for Speech Acoustic Analysis
- Why Speech Acoustic Analysis?
- Annotation
  - Segmentation
  - Alignment
- Speech Analysis and Transcription Software

# Methods for Speech Acoustic Analysis

- Analog-to-Digital Conversion
  - Firstly the sound wave has to be digitized – **sampling** and **quantization**
- Oscillogram analysis
  - Noise, intensity, duration and rhythm analysis
- Spectral analysis
  - FFT, *Fast Fourier Transform*
    - noise and formant structure analysis
  - LPC, *Linear Predictive Coding*
    - formant structure analysis
- Spectrogram analysis
  - Noise, intensity, duration, format structure and rhythm analysis
- Melody analysis
- Intensity analysis

# Why Speech Acoustic Analysis?

- Phonetic description
- Linguistic variability
  - Geographical
  - Social
  - Style
- Speech technologies development
  - **Text-To-Speech systems**
  - **Speech Recognition systems**
  - Dialog systems
- Speech pathology analysis and rehabilitation
- Phonetic conflict analysis for non-native speakers
- Speaker identification

- “The annotation of a speech corpus denotes all symbolic information that is directly related to the speech signal, either via the physical time scale, in which case we speak of a *segmentation and labeling*, or via some semantic content of the speech signal, in which case we speak of a *transcription or tagging*” (Florian Schiel, Christoph Draxler, “The Production of Speech Corpora”, 2004)

# Segmentation vs Alignment

- **Segmentation**
  - It determines the sound or graphic time limits of the units chosen (sentence, word, syllable)
  - Can be automated with reasonable success for good quality recording. A method which detects the more or less abrupt changes in the spectrogram along the time axis (Cosi, 1997)
- **Alignment**
  - It links sound units with corresponding text units
- **Automatic alignment**
  - Markov Models
    - the limits of speech sound are obtained from the phonetic transcription. (Talin and Wightman 1994, Fohr, Mari, et Haton 1996)
  - By synthesis
    - Comparison between the time variations of the speech signal spectra with another speech signal, generated by a text to speech synthesizer operating on the text to align. (Malfrère and Dutoit, 2000)
- **Problems with automatic alignment and segmentation**
  - Performs depends on speaker's voice characteristics
  - Require good quality recordings to reduce the error rate
  - Overlapping of speakers voices

# Speech Analysis and Transcription Software

- [ELAN, Max Plank Institute for Psycholinguistics](#)
- [PCquirer, Scicon R&D, Inc.](#)
- [PitchWorks, Scicon R&D. Inc.](#)
- [Praat, Institute of Phonetic Sciences, University of Amsterdam](#)
- [Prosogram, P. Mertens, Department of Linguistics, KU Leuven](#)
- [SFS, Speech Filing System, Department of Phonetics and Linguistics, University College London](#)
- [Speech Analyzer, CCS Software Development](#)
- [Speech Studio, Laryngograph Ltd.](#)
- [Transana, Wisconsin University](#)
- [Transcriber, C. Barras, LIMSI, CNRS - E. Geoffrois, DGA, CTA, GIP](#)
- [WaveSurfer, Centre for Speech Technology, KTH](#)
- [Winpitch, Pitch Instruments Inc.](#)

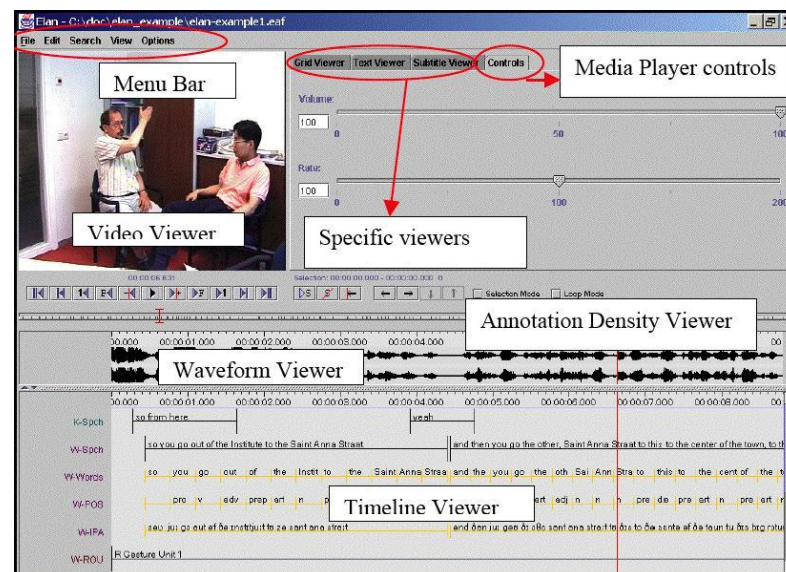


# ELAN, EUDICO Linguistic Annotator

Max Plank Institute for Psycholinguistics

Microsoft | Development Center  
Portugal

- <http://www.mpi.nl/tools/elan.html>
- “It is an annotation tool that allows you to create, edit, visualize and search annotations for video and audio data (...) for purposes of annotation, analysis and documentation.”
  - display a speech and/or video signals, together with their annotations
  - time linking of annotations to media streams
  - linking of annotations to other annotations
  - unlimited number of annotation tiers as defined by the users
  - different character sets
  - export as tab-delimited text files
  - Search options



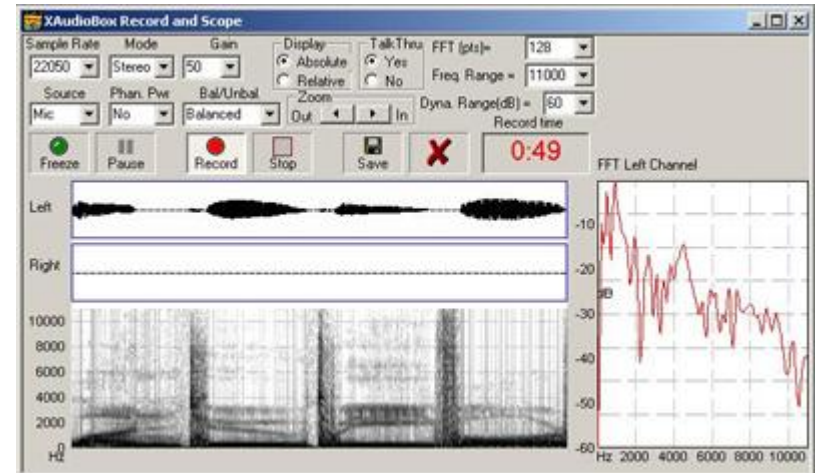


# PCquirer

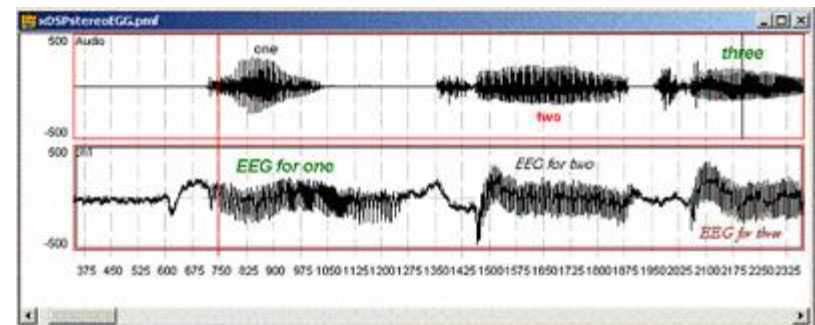
Scicon R&D, Inc.

Microsoft | Development Center  
Portugal

- <http://www.sciconrd.com/pcquirerx.html>
- Operates **XAudioBox** and **XAudioButtonBox** for high quality stereo recording
- PlotFormants function directly from spectrogram or from external file
- Automatic data logging with a click of the mouse
- Real time spectrogram and FFT when recording
- Tape recorder style controls for recording
- Highly controlled playback output level control for perception experiments
- Stereo operation with sample rate of 11 , 22 and 44 kHz
- Sample rate, gain, filter rate controlled by software
- Reads many audio data types



Real time spectrogram and FFT of audio



Free hand labeling capability

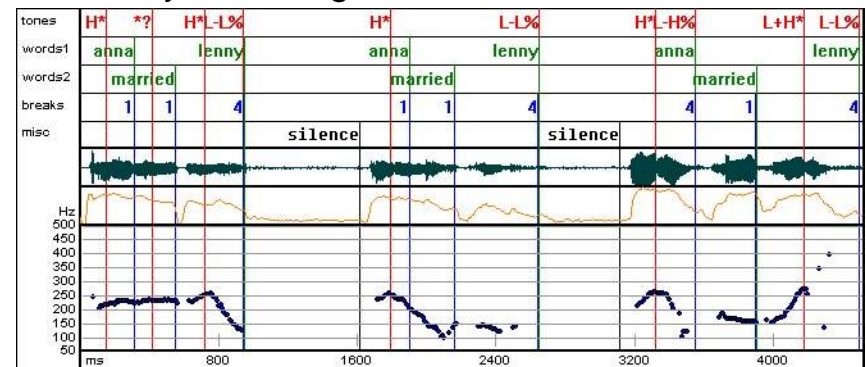
# PitchWorks

Scicon R&D

Microsoft Development Center  
Portugal

- <http://www.sciconrd.com/pitchworks.html>
- “**PitchWorks** is the main tool for any intonation studies.”
- 10 levels of tiers
- **TOBI** style labeling
- Capable of reading many different file types.
- FFT, LPC, Intensity, Spectrogram, Formant tracking, ...
- Cepstral and Autocorrelation, pitch extraction methods
- Synchronized cursor between windows
- Automatic data logging
- Direct printing from every window
- Save each window as a bitmap (PC)

## TOBI style labeling



View of the label window - The labels can be sorted by tiers, labels, or time. A label can be selected and the file can be zoomed to that label. No need to look into a long file to look for any labels

Labels: GoNorth

Click on the column header to sort.      Zoom to Selected Label

	Tier	Label	Time(ms)
1	words	Aahm.	399.68
2	words	just	632.53
3	words	go	775.03
4	words	north	1,139.95
5	words	from	1,299.83
6	words	there	1,595.24
7	words	through	1,831.53
8	words	the	1,890.67
9	words	Tech	2,220.82
10	words	Center	2,582.29

# Praat

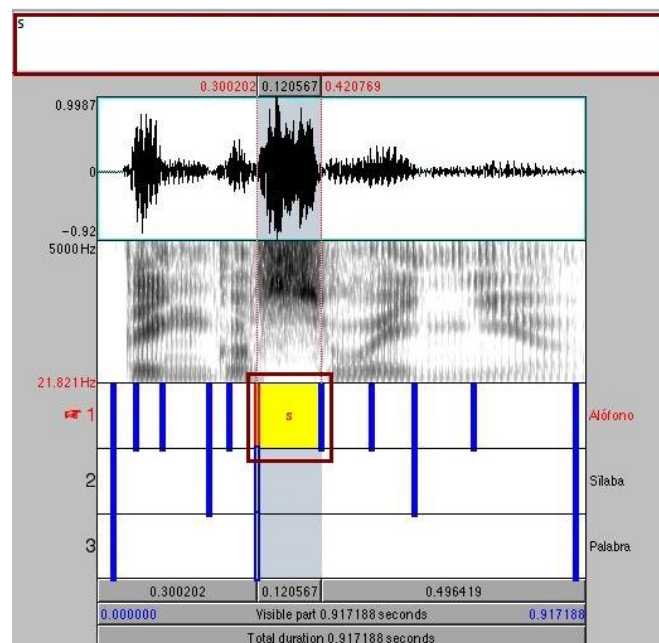
P. Boersma & D. Weenink,

Institute of Phonetic Sciences University of Amsterdam

Microsoft

Development Center  
Portugal

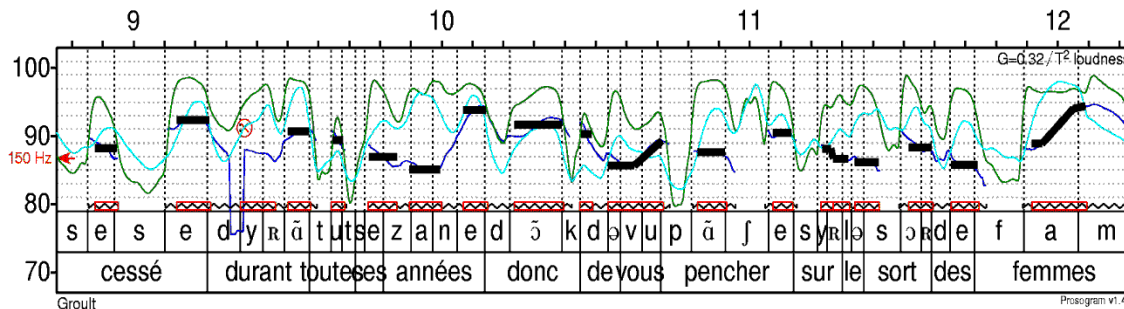
- <http://www.praat.org>
- Developed by Paul Boersma and David Weenink, in Dutch means “Talk”
- General purpose speech tool :
  - Speech analysis
  - Speech synthesis
  - Segmentation and labeling
  - Speech manipulation
  - Learning algorithms
  - Statistics
  - Listening experiments
- Online help, FAQ, manual
- [Additional tutorials, scripts, resources, user groups](#)



# Prosogram

P. Mertens, Department of Linguistics, KU Leuven

- <http://bach.arts.kuleuven.be/pmertens/prosogram/>
- “Transcription of prosody using pitch contour stylization based on a tonal perception model and automatic segmentation (...)”
  - F0, intensity, voicing (V/UV)
  - Obtain a segmentation
  - Stylize the F0 of the selected time intervals
  - Determine pitch range used in speech fragment. Plot stylized pitch and some annotation tiers (text, phonetic transcription).
  - Use a musical (semitone) scale and add calibration lines at every 2 ST for easy interpretation of pitch intervals.
- The system is implemented as a [Praat](#) script



## Speech Filing System, University College London

- <http://www.phon.ucl.ac.uk/resource/sfs/>
- Tutorials : <http://www.phon.ucl.ac.uk/resource/sfs/howto/>
- “(...) It performs standard operations such as acquisition, replay, display and labeling, spectrographic and formant analysis and fundamental frequency estimation. It comes with a large body of ready made tools for signal processing, synthesis and recognition, as well as support for your own software development.”
- Acquisition and replay
- Waveform processing, Laryngographic processing
- Fundamental frequency estimation formant frequency estimation
- Formant synthesis
- Spectrographic analysis
- Filterbank analysis/synthesis
- Resampling
- Speed/pitch changing
- Annotation
- Spectral cross-sections waveform envelope
- Filtering
- Signal editing
- Signal alignment

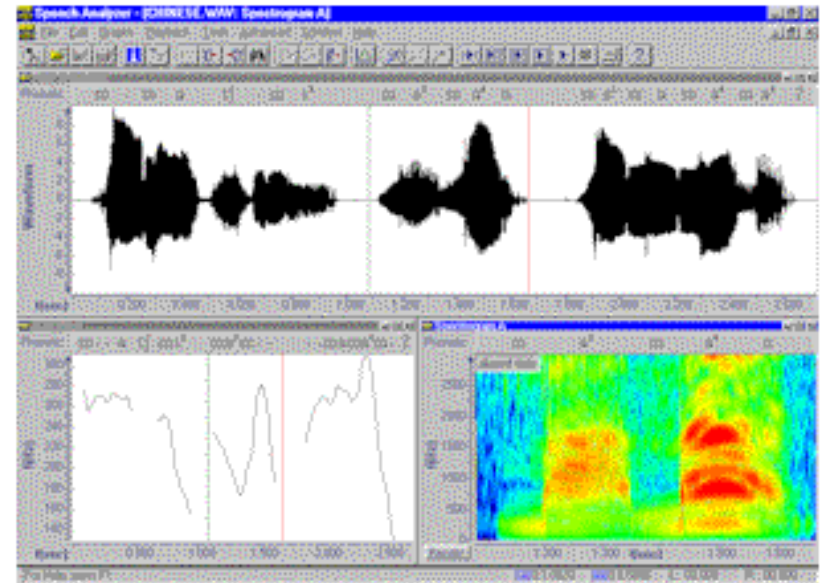


# Speech Analyzer

CCS Software Development

Microsoft | Development Center  
Portugal

- <http://www.sil.org/computing/speechtools/speechanalyzer.htm>
- “Use this software for recording, transcribing, and analyzing speech files.”
  - Transcribe speech files phonetically with IPA.
  - Playback at a slower speed.
  - Playback with repetition with variable length delay between repetitions.
  - Add phonemic, orthographic, tone and gloss annotations to your transcription in an interlinear format.
  - View sound file as a waveform, pitch plot, spectrogram, spectrum and various F1 vs. F2 displays.
  - Music Analysis capability



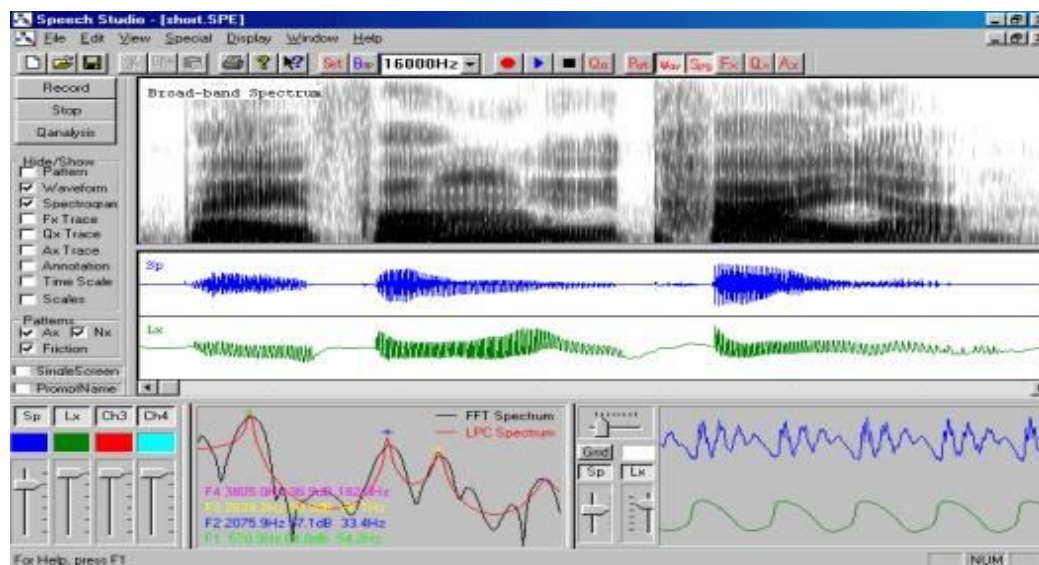


# SpeechStudio

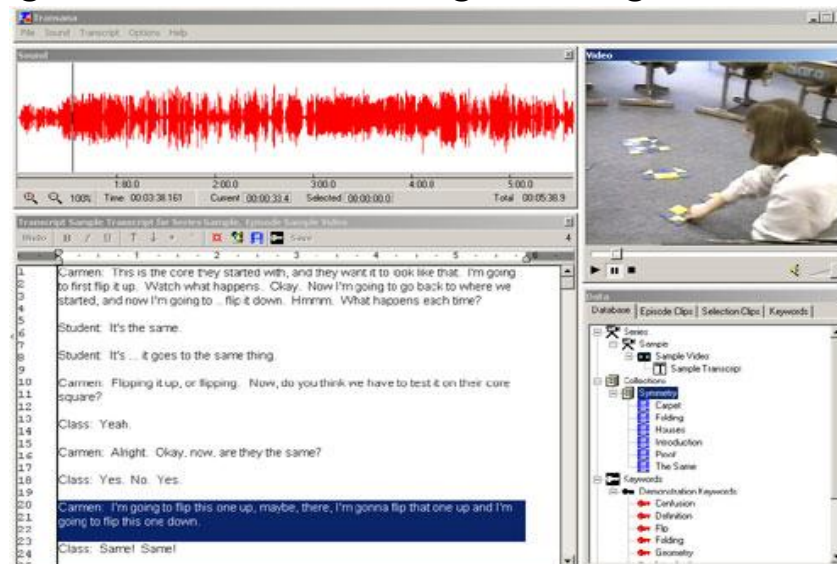
Laryngograph Ltd.

Microsoft | Development Center  
Portugal

- [http://www.laryngograph.com/pr\\_studio.htm](http://www.laryngograph.com/pr_studio.htm)
- “Speech Studio is a software and hardware package, which has been specially designed for phoneticians, speech scientists and quantitative work by ENT clinicians and SLT’s. It supports data recording direct to hard disk, real-time displays, and instantaneous quantitative analysis and pattern target mode for speech training.”



- <http://www.transana.org/>
- “Transana is designed to facilitate the transcription and qualitative analysis of video and audio data. It provides a way to view video or play audio recordings, create a transcript, and link places in the transcript to frames in the video (...) It also features database and file manipulation tools that facilitate the organization and storage of large collections of digitized video.”



# Transcriber

C. Barras, LIMSI, CNRS - E. Geoffrois, DGA, CTA, GIP

Microsoft | Development Center  
Portugal

- <http://www ldc.upenn.edu/mirror/Transcriber/>
- “Transcriber is a tool for assisting the manual annotation of speech signals”
  - user-friendly graphical user interface
  - speech recordings
  - segmentation
  - Transcription
  - Labeling
- It is more specifically designed for:
  - annotation of broadcast news recordings
  - creating corpora used in the development of automatic broadcast news transcription systems



# WaveSurfer

Centre for Speech Technology, KTH

**Microsoft** | Development Center  
Portugal

- <http://www.speech.kth.se/wavesurfer/>
- “WaveSurfer is an Open Source tool for sound visualization and manipulation ”
  - Flexible interface - handles multiple sounds
  - Common sound file formats - reads, and writes WAV, AU, AIFF, MP3, CSL, SD, Ogg/Vorbis, and NIST/Sphere
  - Transcription file formats - reads, and writes HTK (and MLF), TIMIT, ESPS/Waves+ and Phondat. Support for encodings and Unicode
  - Unlimited file size - playback and recording directly from/to disk
  - Sound analysis - e.g. spectrogram and pitch analysis
  - Customizable - users can create their own configurations. Localization support
  - Extensible - new functionality can be added through a plugin architecture
  - Embeddable - WaveSurfer can be used as a widget in custom applications
  - Scriptable - hosts a built-in script interpreter

# Winpitch

Pitch Instruments Inc.

Microsoft | Development Center  
Portugal

- <http://www.winpitch.com>
- Multimedia
- Real time monitoring of recordings (spectrogram, Fo, etc.)
- High precision segmentation, speech turns overlapping
- Direct transcription capability
- Assisted alignment of existing transcription
- Automatic building of speech segments database (XML output)
- Multimedia input formats (wav, mp2, aiff, au, mpeg, mp2, mp4, avi, gsm 6.1, etc.)
- Direct speech analysis (spectrogram, Fo, intensity, etc.) from speech segments database
- Prosodic morphing

