INTERSPEECH 2019: Show & Tell Contribution
September 15–19, 2019, Graz, Austria

Sound Tools eXtended (STx) 5.0 – a powerful sound analysis tool optimized for speech

Anton Noll, Jonathan Stuefer, Nicola Klingler, Hannah Leykum, Carina Lozo, Jan Luttenberger, Michael Pucher, Carolin Schmid

Acoustics Research Institute, Austrian Academy of Sciences, Vienna

[anton.noll|jonathan.stuefer|nicola.klingler|hannah.leykum|carina.lozo]@oeaw.ac.at,
[jan.luttenberger|michael.pucher|carolin.schmid]@oeaw.ac.at

Abstract

In this paper, we introduce Sound Tools eXtended (STx) version 5.0, an acoustic speech and sound processing application. STx 5.0 contains an integrated, simplified and compact GUI, specifically designed for speech analysis for phoneticians, linguists, psychologists, and researchers in related fields. It features a well structured user interface, compatibility with established tools (TextGrid [1], MAUS [2]), and top-notch signal analysis tools. STx 5.0 enables researchers as well as students to conduct advanced analysis of audio files, especially of speech recordings. STx 5.0 implements a new interface for the already established profiles in STx 5.0, which helps customize settings according to the researcher's needs.

Index Terms: speech analysis, speech analysis software, phonetics

1. Introduction

Not only phoneticians, but also (applied) linguists, philologists, and scholars of related disciplines strive towards data driven research. There is growing demand for empirical speech data collection and analysis in the field of humanities. However, only few researchers in the humanities learn in their academic training how to use signal processing software. Furthermore, and especially while working with data which is not recorded in sound attenuated booths (e.g. dialectological or typological field work), researchers might face limitations regarding the software: Field work is often recorded as one file, with one or many people speaking simultaneously (experimentee and experimenter). Therefore, the recordings have to be structured manually. Moreover, many automatic formant trackers produce contorted output when used with default parameters (e.g. not adjusting for the age or gender of the speaker [3]) which is also a problem while working with a variety of different speakers. Accessing the segmented and analyzed data and further processing it for statistical analyses can also be a huge obstacle in experimental designs.

In this paper we will present STx 5.0, to demonstrate the new easy-to-use workspace mode and other advantages of STx 5.0.

1.1. Sound Tools eXtended 5.0

STx 5.0 is the newest version of STx [4] which is downloadable for all prevalent Microsoft Windows distributions and Wine for free (*https://www.kfs.oeaw.ac.at/stx*). Due to its very specific applications, and close exchange between the programmers and scientists, previous versions of STx were already a powerful tool with a lot of advanced features. To ease the first steps towards speech analysis for new users, or non-technical researchers (like students of linguistics, philology, logopeadics, ...) STx 5.0 has been developed in collaboration with researchers from the field of acoustic phonetics.

The new workspace mode was particularly developed for speech analysis, in which the complex features unrelated to speech analysis were hidden in order to have a simplified user interface which facilitates the use of the software.

1.2. STx application domains

STx has already been used in many scientific and technical fields from phonetics to bioacoustics, due to its huge variety of signal processing features e.g. frequency analysis (Fast Fourier Transformation, Discrete Fourier Transform, wavelet, filter banks), spectral smoothing, de-noising, and signal enhancement. It has been used in *phonetics* for investigations of gender differences [5], phonotactics [6], and the spread of Viennese laterals [7], in *speech synthesis* for creation of resources for dialect synthesis [8], in *forensic speaker recognition* for the analysis of disguised voices [9], the influence of coding on formant tracking [10], and the effect of recording distance and direction on formants [11], in *bioacoustics* for analysis of sound and vocalisations of fish [12], elephants [13], and mice [14], in the *analysis of noise signals* for analysis of train rolling noises [15], as well as in *musicology* [16], and *cognitive neuroscience* [17].

2. Simplifying Speech Analysis with STx 5.0

Since until now, several features of STx were only usable with programming knowledge or advanced computer skills, STx 5.0 aims to ease the work needed for speech analysis. In the present paper we will list and explain the most relevant features in order to show how one can analyze a signal.

The most important work steps of speech analysis can be done using STx 5.0: With the included recorder, collecting data is possible; sound files can be stored in a non compressed format; acoustic characteristics can be analyzed, and measurements can be summarized by descriptive statistics; in addition, a real time analyzer allows the live demonstrations of acoustic properties.

STx 5.0 allows for an easy workflow, starting with the drag-and-drop import of one or many sound files in the workspace. The sound files can be organized in projects, helping to structure the data while working on the analyses.

STx 5.0 is very suitable for analyzing a large quantity of data. Especially, the segmentation as well as the analysis of long sound files is easy. Equally, a larger number of sound files poses no problems, thanks to the possibility of grouping them

in projects: the user can continue working without the need to manually open all sound files again. Likewise, customizable settings can be saved in the profile to ensure consistent settings in following analysis sessions.

STx 5.0 will have a call function for WebMAUS [2, 18] implemented, a web application for automatic segmentation of speech recordings. In combination with the automatic detection of pauses/silence and the possibility to import (and edit) word lists or sentence lists, the implementation of WebMAUS constitutes a simple possibility to automatically segment long sound files. For manual segmentation, the transcription tool of STx 5.0 helps define utterance boundaries. For more detailed segmentation, the analysis window provides a combined view of waveform, spectrogram and short time spectrum. A broad set of (customizable) hotkeys facilitates workflow efficiency.

With STx 5.0 it is easy to follow data protection regulations: during segmentation, personal or sensitive data can be marked to be anonymized. When running the anonymization tool, the selected data can be either replaced by silence, white noise, a 440 Hz tone, or a 1000 Hz tone, simultaneously, transcribed text can be replaced by a replacement text (e.g. <anonymized>).

The workspace of STx 5.0 is partially hierarchically structured. Segments are directly linked to the sound file, the number of subsegments is not limited, allowing different levels of segmentation (e.g.: sentence, syllable, word, phoneme, phones). In combination with the comprehensive find dialog (hotkey: ctrl+f), the extraction of parameters or the export of segments is simple.

The user-friendly possibility to manually correct the measurements (formants, f0, intensity contour) enables the analysis of sound files which were recorded in non-perfect conditions and also where automatic tracking is inaccurate.

In order to gain an overview of the data, basis statistical analyses (mean, standard deviation, duration, min, max, median, variance) can be reported directly in STx 5.0. For more elaborated analyses, the parameters can be extracted and exported in a table (.csv, .xls, .xlsx, .txt).

Not only the export of segments as individual sound files, but also the export of the segment boundaries and names as a TextGrid file is possible. Therefore, the user has the option to shift to other speech analysis software (like PRAAT [1]). It works the other way round as well: the information contained in TextGrid files can be imported into STx 5.0.

For additional information regarding STx see *https://www.kfs.oeaw.ac.at/stx* for the documentation and additional features.

3. Summary

We have presented Sound Tools eXtended (STx) 5.0, the newest version of STx which supports speech analysis scientists by providing a top-notch signal analysis tool (downloadable from the website *https://www.kfs.oeaw.ac.at/stx*). Its various features which were developed in close cooperation with trained phoneticians and (applied) linguists ensure an intuitive, scientifically approved workflow to enable young researchers as well as researchers without advanced knowledge of signal analysis to process speech data reliably and robustly. Since STx 5.0 is aimed to be constantly improved, we welcome all feedback and improvement ideas.

4. References

[1] P. Boersma and D. Weenink, "Praat: doing phonetics by computer," Amsterdam, 2017. [Online]. Available: http://www.praat.org

[2] F. Schiel, "Automatic phonetic transcription of non-prompted speech," in *Proceedings of the ICPhS*, 1999, pp. 607–610.

[3] F. Schiel and T. Zitzelsberger, "Evaluation of automatic formant trackers," in *Proceedings of the Eleventh International Conference on Language Resources and Evaluation, LREC 2018, Miyazaki, Japan, May 7-12, 2018*, 2018.

[4] P. Balazs, A. Noll, W. Deutsch, and B. Laback, "Concept of the integrated signal analysis software system STx," in *Jahrestagung der Österreichischen Physikalischen Gesellschaft 2000, ÖPG 2000*, 2000.

[5] C. Schmid and S. Moosmüller, "Gender differences in the phonetic realization of semantic focus," in *Proceedings of Prosody-Discourse Interface (IDP 2013)*, Leuven, Belgium, 2013, pp. 119–123.

[6] H. Leykum and S. Moosmüller, "Phonotaktische und morphonotaktische Konsonantencluster in wortmedialer Position in der österreichischen Standardaussprache," in *Dimensionen des sprachlichen Raums. Variation – Mehrsprachigkeit – Konzeptualisierung*. Peter Lang, 2019, pp. 127–145.

[7] M. Rausch-Supola, S. Moosmüller, C. Schmid, and H. Leykum, "Die Ausbreitung des Wiener velarisierten Laterals: ein Vergleich Wien - Neunkirchen." Graz, Austria: 42. Österreichische Linguistiktagung, 2016.

[8] M. Pucher, F. Neubarth, V. Strom, S. Moosmüller, G. Hofer, K. C., G. Schuchmann, and S. D., "Resources for speech synthesis of Viennese varieties." in *Proceedings of the 7th International Conference on Lanquage Resources and Evalution (LREC)*, Valleta, 2010, pp. CD–ROM.

[9] E. B. Brixen and E. B. Brixen, "Digitally disguised voices," in *Audio Engineering Society Conference: 39th International Conference: Audio Forensics: Practices and Challenges*, June 2010.

[10] E. Enzinger, "Measuring the effects of adaptive multirate (AMR) codecs on formant tracker performance." *Acoustical Society of America Journal*, vol. 128, p. 2394, 2010.

[11] E. B. Brixen and S. Christensen, "Influence of recording distance and direction on the analysis of voice formants - initial considerations," in *Audio Engineering Society Convention 131*, Oct 2011. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=16021

[12] F. Ladich, "Females whisper briefly during sex: context-and sex-specific differences in sounds made by croaking gouramis," *Animal Behaviour*, vol. 73, no. 2, pp. 379–387, 2007.

[13] A. S. Stoeger and A. Baotic, "Information content and acoustic structure of male african elephant social rumbles," *Scientific reports*, vol. 6, pp. 275–285, 2016.

[14] S. M. Zala, D. Reitschmidt, A. Noll, P. Balazs, and D. Penn, "Automatic mouse ultrasound detector (a-mud): A new tool for processing rodent vocalizations," *PLOS ONE*, vol. 12(7), p. e0181200, 2017.

[15] C. H. Kasess, A. Noll, P. Majdak, and H. Waubke, "Effect of train type on annoyance and acoustic features of the rolling noise," *Journal of the Acoustical Society of America*, vol. 134 (2), pp. 1071–1081, 2013.

[16] G. Adamo, "Social roles, group dynamics and sound structure in multipart vocal performance: the female repertoire for the good friday at cassano allo ionio (south italy)," in *European Voices I. Multipart Singing in the Balkans and the Mediterranean*. Wien: Böhlau, 2008, vol. 1, pp. 87–101.

[17] C. Lamm, C. D. Batson, and J. Decety, "The neural substrate of human empathy: Effects of perspective-taking and cognitive appraisal," *Journal of Cognitive Neuroscience*, vol. 19, no. 1, pp. 42–58, 2007.

[18] T. Kiesler, U. D. Reichel, and F. Schiel, "Multilingual processing of speech via web services," *Computer Speech & Language*, vol. 45, pp. 326–347, 2017.