SoftwareX 23 (2023) 101492

Contents lists available at ScienceDirect

SoftwareX

journal homepage: www.elsevier.com/locate/softx

Original software publication

xkl: A legacy software for detailed acoustic analysis of speech made modern

Luca De Nardis [a,*], Maria-Gabriella Di Benedetto [a,b], Jeung-Yoon Choi [b], Stefanie Shattuck-Hufnagel [b]

[a] DIET Department, Sapienza University of Rome, Rome, Italy
[b] Massachusetts Institute of Technology (MIT), Cambridge, MA, USA

ARTICLE INFO

Article history:
Received 27 March 2023
Received in revised form 4 July 2023
Accepted 31 July 2023

Keywords:
Speech analysis
Speech spectral analysis
Spectral display

ABSTRACT

The determination of the fundamental properties of speech relies on a fine and precise estimation of temporal and spectral properties of speech segments. Given the time-varying nature of speech, a digital estimation of its instantaneous spectrum is particularly challenging, and has been the object of investigation throughout the past 50 years. The xkl software, developed in the 80's by the late Dennis Klatt at MIT, has superior capabilities in addressing the above question. Its use in the past 20 years was, however, limited by a lack of support for modern computing platforms. Revamping it will give access to a powerful tool that will eventually lead to new discoveries.

Code metadata

Current code version	v3.2
Permanent link to code/repository used for this code version	https://github.com/ElsevierSoftwareX/SOFTX-D-23-00128
Code Ocean compute capsule	N/A
Legal Code License	GPL v2.0
Code versioning system used	none
Software code languages, tools, and services used	C language, bash scripting
Compilation requirements, operating environments & dependencies	Operating Systems: Linux, macOs, Windows. Requirements: openmotif, xorg and sox packages (Linux) homebrew and its motif, xorg and sox packages (macOs), cygwin and its motif, xorg and sox packages (Windows)
Link to developer documentation/manual	http://acts.ing.uniroma1.it/xkl.php
Support email for questions	xkl@uniroma1.it

1. Motivation and significance

Processing speech signals and analyzing their temporal and spectral properties is crucial in speech science. Research investigations of all sorts in fields such as, for instance, automatic speech recognition [1–4], speech disorders [5–7], second and child language acquisition [8], and socio-linguistics [9], require fine and precise measurements of the acoustic signal. Given the time-varying nature of speech, a digital estimation of its instantaneous spectrum is particularly challenging and has been the object of investigation for the past 50 years. Throughout the years and decades, the development of speech processing software has therefore been a thriving activity in supporting research in this field. An example of popular software in the community is Praat [10], developed by Paul Boersma and David Weenink, that runs under multiple operating systems including Windows, macOs and Linux. Other software packages include Wavesurfer [11], and VoiceSauce [12] that is based on Matlab. The xkl software, developed in the 80's by the late Dennis Klatt at MIT, has outstanding capabilities in providing reliable and user friendly tools for extracting signal measurements [13,14]. This paper describes some of the capacities of this software, and its updating to run on current platforms.

When xkl first became available, in the 1980s, its capacity for detailed representation of the linguistically-relevant aspects of

* Corresponding author.
E-mail address: luca.denardis@uniroma1.it (Luca De Nardis).

https://doi.org/10.1016/j.softx.2023.101492

the acoustic speech signal was revolutionary. The result was that a variety of researchers began to use it, and adapt it to their local computer environments. The impact of Klatt's synthesis module, called KlSyn, which is included in the set of tools provided by xkl, is evident for it being ported to C about 30 years ago [15] and then released in several other languages over the years, such as Python [16] and Java Script [17]: as of today, a C++17-language version of Klatt's synthesizer is integrated in Praat, making it available on modern operating systems [10]. However, the same did not happen for xkl as a whole, and more specifically its spectral analysis modules, because its source code was never released to the wider audience.

The subsequent fate of xkl reflects two developments that caused this software to be used less widely. The first was a shift in emphasis from the goal of understanding what information might be available in the signal for human listeners to use, to the goal of maximizing performance of Automatic Speech Recognition (ASR) systems. This shift resulted in a downplaying of the value of explicit analysis of the individual cues to phonological contrasts, and of how those cues and cue values vary systematically across speakers, contexts and dialects. The second development was an increasing recognition of the power of machine learning to exploit detailed acoustic pattern information without making this information available to investigators. The resulting explosion of ML-based ASR systems with practical usability also downplayed the value of explicit cue analysis.

In recent years, however, the field has begun to recognize the value of an explicit understanding of which types of information are most useful to extract from the signal for speech recognition [18]. We thus believe that the time is ripe for updating the xkl software to run conveniently on computers using current operating systems. This tool will enable a detailed level of acoustic analysis that reveals the systematic differences among phonological categories, contextual realizations of those categories, speaker populations and individual speakers, that are increasingly recognized as critical information, not only for the development of ASR systems that work well for various populations, but also for understanding how human speakers and listeners process speech signals.

Updating the xkl tool to run efficiently and reliably on current operating systems will, we believe, provide the resource that the field is looking for. Several recent papers have illustrated the usefulness of this analysis tool [19–24]. We envision that a more accessible and convenient version of the xkl tool may lead to many important discoveries about the characteristics of the acoustic speech signal that speakers and listeners control, attend to and interpret. This paper illustrates the first step in this direction: reviewing and updating xkl source code and documentation to make it usable on modern platforms. A description of further evolution planned for the xkl tool is provided in the concluding discussion.

Providing speech samples for use in xkl is straightforward; the user can either collect speech data by recording it directly within xkl, or upload pre-recorded speech materials. All the processing functions are available within a unified user interface, that is described in the next section.

2. Software description

xkl provides unified access to a comprehensive set of tools for reproduction, analysis and synthesis of speech signals, developed by Dennis Klatt as independent command line programs, through a Graphical User Interface (GUI) based on the Motif open source software libraries [25]. Our proposed xkl version (3.2) is developed in C, as was the original version, but now runs under Windows, Linux and macOs. Most of the files in the original code were modified in order to allow the software to build correctly using modern versions of gcc and clang compilers and the latest available version of the OpenMotif libraries; as indicated in Section 1 the focus in this release was not on adding features, but rather on reviving the software. This goal was also pursued by ensuring that the software source code is openly available to the research community: xkl is in fact made available through a GPL 2.0 open source license; license exemptions are also available upon request. Please refer to [26] for further information.

2.1. Software architecture

The xkl source code consists in a set of C files organized in five folders:

- utils, including files related to utilities for file format conversion to and from the Klatt .wav format used by the software;
- common, including files related to playing/recording waveforms from/to .wav files and reading/writing them;
- syn, including files related to the Klatt synthesizer KlSyn;
- lspecto, including files related to the Klatt spectrogram analysis tool lspecto;
- xkl, including files related to the GUI and its interaction with functions provided by files in other folders.

The source code structure of xkl is presented in Fig. 1. The figure shows the C files in each folder; files represented as orange rectangles correspond to executables once compiled, while files represented as green rounded rectangles are support C files used by other files and by executables. Arrows highlight dependencies; an arrow from a file A to a file B indicates that A depends on functions defined in B. Fig. 1 highlights that the software provides executables in all but the common folder. Executables in the utils folder provide a set of utilities for .wav files management, that allow the user to extract data from .wav files, play them, and convert them to and from other formats. The synmain executable in the syn folder is the command line version of the KlSyn synthesizer, that can run independently from the GUI, if needed; the version provided is v.2.1 of KlSyn93. Similarly, the lsp executable is the command line version of the lspecto tool to compute and print spectrograms for a waveform stored in a .wav file.

Functionalities provided by the command line tools synmain and lsp are fully integrated in the xkl executable (found in the xkl folder), that is the main focus of this work. The functionalities provided by xkl are described in the following subsection.

2.2. Software functionalities

The xkl executable loads the GUI and provides five dropdown menus giving access to five sets of functions:

- File Input/Output — functions related to reading and writing files;
- Time analysis — functions related to waveform graphical representation and time-domain analysis;
- Frequency analysis — functions related to spectrum computation and graphical representation;
- Reproduction and recording — functions related to reproduction and recording of speech files;
- Synthesis — functions related to speech synthesis.

The menus and corresponding sets of functions are described below.

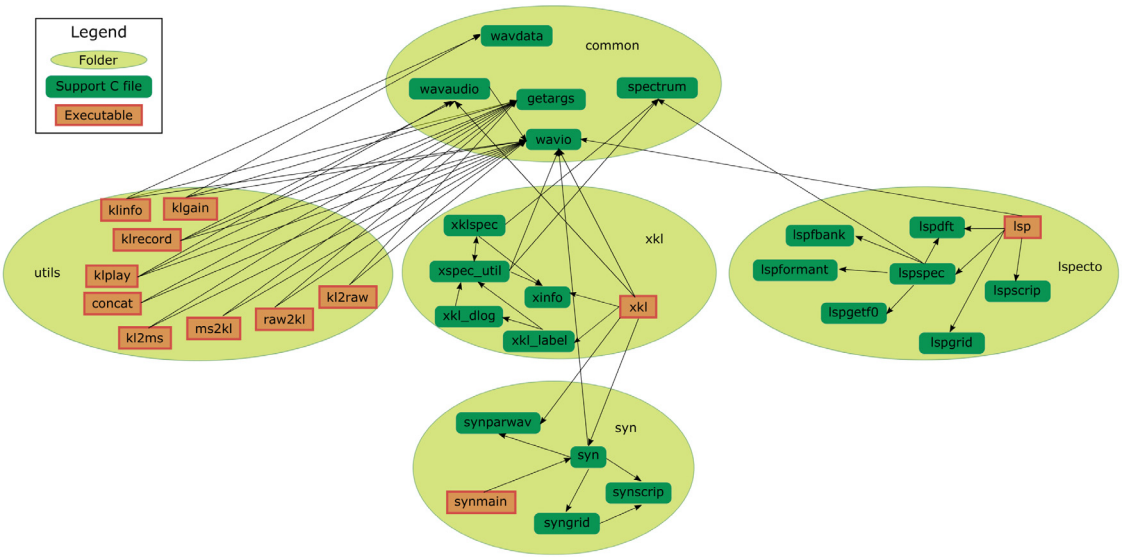

Fig. 1. Source code structure of the xkl software. An arrow indicates that the code in the start file depends on functions defined in the end file; orange rectangles indicate files that correspond to executables, while rounded green rectangles indicate support C files used by other files and by executables.

2.2.1. File input/output

The main function provided in this menu is reading/writing speech recordings. xkl originally encoded and decoded speech files according a proprietary format called Klatt wave (.wav), consisting of a custom header followed by signal samples stored as 16-bit unsigned integers. The current release maintains support for the original Klatt format; the most common and widely used Microsoft .wav format, employed by most recording tools, is however supported by the conversion tools in the utils folder.

In addition to speech files, xkl supports text label files, postscript files used to print time and frequency analysis plots, and ASCII text files for import/export of time and frequency data related to a waveform.

2.2.2. Time analysis

This menu includes functions related to graphical representation and handling of the signal waveform, including time shift, zoom in and out, and cursor setting. It also provides options to enable time-domain analysis, in particular peak/valley picking.

2.2.3. Frequency analysis

This menu includes functions related to the representation and analysis of the signal in the frequency domain, including Digital Fourier Transform (DFT) with a resolution of up to 4096 points, and smoothed and Linear Prediction Coding (LPC) spectra as well as spectrograms. The analysis functions also include speech-specific options such as peak picking, formant estimation, and critical band computation.

2.2.4. Reproduction and recording

This menu includes functions related to playing and recording audio files. The entire content of the waveform currently displayed can be played out; it is also possible to select and play back a specific time interval. Recording speech material is possible by choosing the "Record as .wav file" entry from the menu. Before doing so, the menu also offers the possibility of setting recording parameters, that is sampling rate and predetermined recording duration. The recorded waveform is saved on disk in both Klatt and Microsoft .wav formats, and automatically displayed in xkl for analysis.

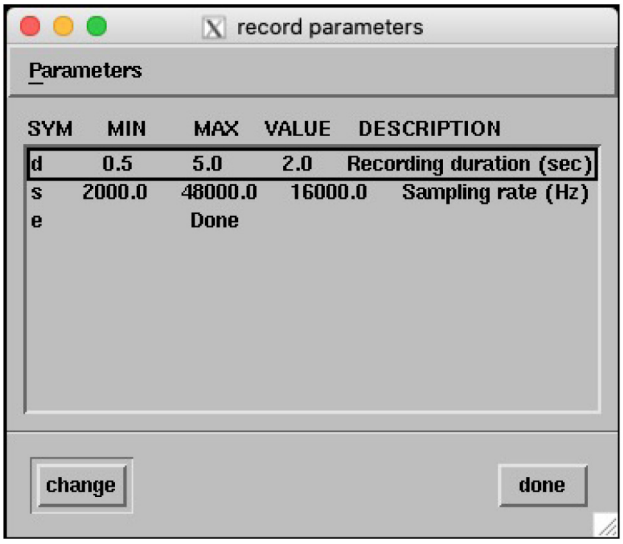

Fig. 2. Dialogue window for recording settings in xkl.

2.2.5. Synthesis

This menu provides access to the KlSyn formant synthesizer [27]. Upon running the "Synthesize" command xkl accepts a .doc file containing formant frequency data vs. time and other constant vs. variable synthesizer control parameters, to then produce a synthetic waveform file. The formant synthesizer is described in [27], except for the existence in xkl of the possibility of selecting a more natural glottal source waveform.

3. Illustrative examples

The software xkl is a superior tool when it comes to performing acoustic analyses of speech. In particular, it provides extremely high-quality spectrograms and the capacity to extract, plot, and analyze in a very straightforward manner the spectrum of a particular signal window (spectrum slice) at any instant of time along the time axis. These features are relevant for any study that requires measuring – with high precision – acoustic parameters in the frequency domain. This is particularly true for

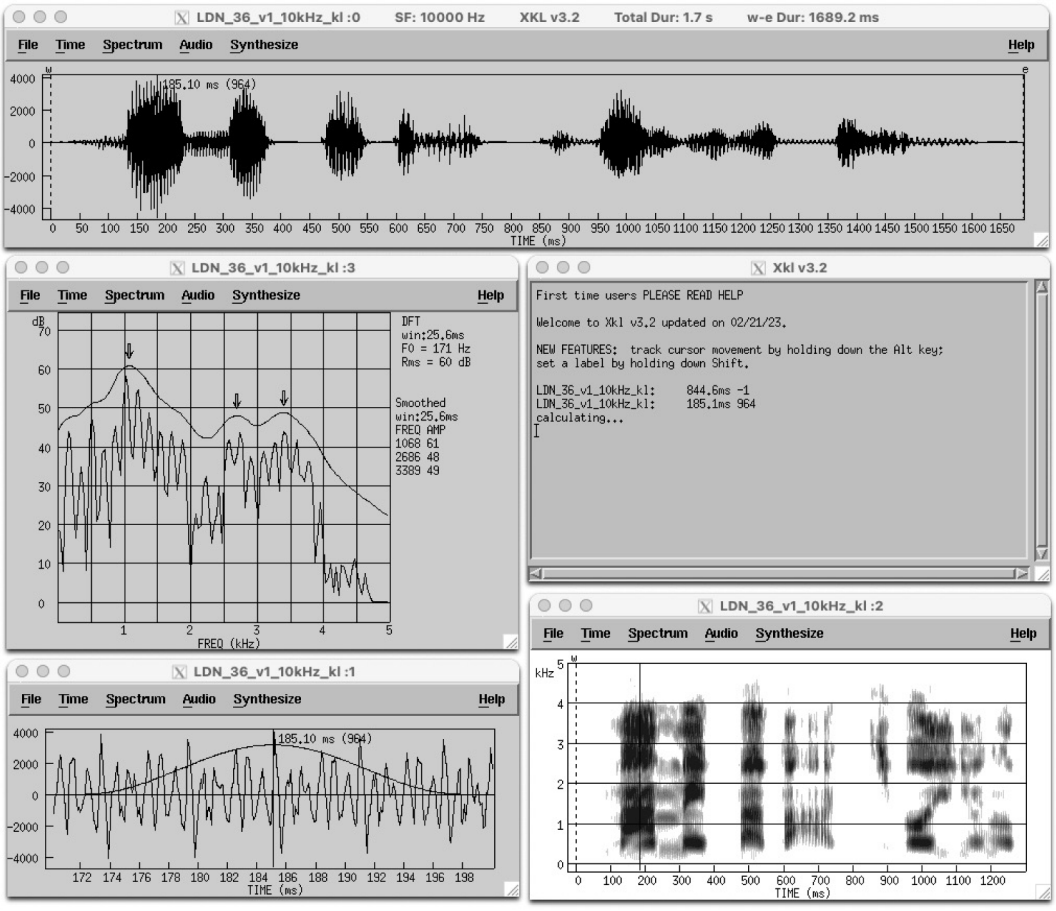

Fig. 3. Xkl windows related to a speech signal and shared feedback window. Windows related to the speech signal are numbered with a suffix :x after the filename in the window header, with *x* from 0 to 3. Window 0: signal waveform; Window 1: zoom on waveform section around cursor position and shape of the window used to select the portion used for spectral analysis; Window 2: signal spectrogram; Window 3: spectrum slice.

the analysis of vowels, where the need for estimating formant frequencies with a precision of a few Hz may be real, as in studies of rounding [4,28,29].

Two illustrative examples of the use of xkl are provided below. The first shows how to acquire or import a speech signal, how to observe and analyze it in both time and frequency domains, and finally how to export a vectorial image of one or more spectrum slices and/or spectrogram. The second example focuses on the use of xkl to measure formants in a challenging signal, i.e. speech produced by a young female speaker with high pitch.

3.1. Speech sample acquisition and analysis

xkl allows the recording of mono signals with a sampling rate in the range [2 kHz–48 kHz]. Fig. 2 shows the dialogue window for the recording parameters setup. Note that xkl operates with a predefined recording duration that can be configured in this window, along with the sampling rate.

xkl also supports importing .wav audio files recorded with external tools, as discussed in Section 2.2. Following a recording or an import operation, xkl displays four windows allowing the analysis of the signal in both time and frequency; xkl also shows a window providing feedback on the required operations, shared by all open speech signals. Such windows are shown in Fig. 3 for the example sentence, in the Italian language, "Mamma e Papà ti vogliono bene" ("Mom and Dad love you"), corresponding to sentence number 36 of the LaMIT database [21,30,31], created as part of the LaMIT project [32], in which the lexical access model developed by Stevens in [33] is applied to Italian. Windows related to the speech signal are numbered with a suffix ":x" after the filename in the window header, with *x* from 0 to 3. Window 0 shows the waveform, and the position of the cursor in this window determines the content shown in the other windows. Window 1 shows the segment of the signal selected by the windowing function around the cursor position, and the windowing function itself, while window 3 contains the spectrum slice of the selected signal segment. Window 2 shows the spectrogram.

Each window can be exported in a postscript file; xkl also allows to export up to 4 different spectrum slices in the same postscript file, making the comparison between different slices extremely convenient. It is worth noting that xkl allows the user to open multiple speech signals at the same time, and to export slices from different signals into the same postscript file. An example of a postscript image containing 4 spectrum slices from two different files is presented in Fig. 4.

3.2. Formant analysis

As illustrated above, a valuable feature of xkl is its capacity to provide high-quality spectrum analysis and spectrograms. Moreover, xkl also computes up to the fifth formant for vowel segments according to an algorithm conceived by Dennis Klatt that is capable of estimating formants with remarkable precision. In this section, we present an example of analysis of the vowel [i] belonging to the Italian bysillabic sequence [izi] pronounced by a young female speaker of age 13. This case was selected on

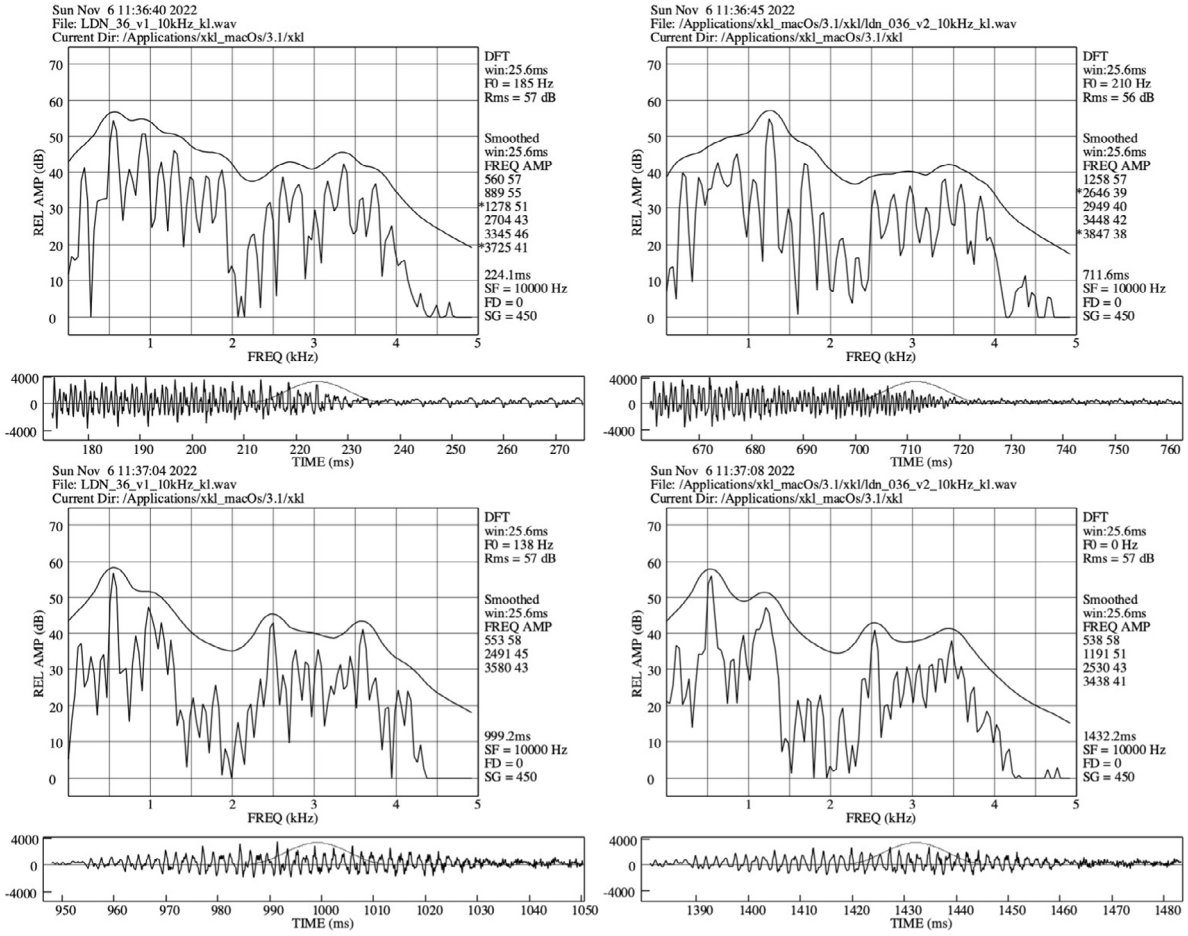

Fig. 4. Example of postscript image containing 4 spectrum slices, collected at different time points from two different recordings of sentence 36 of the LaMIT database.

purpose, to illustrate the complexity of analyzing high-pitched voices in young adults. In order to highlight the possibilities offered by xkl, we compared xkl formant values with manually extracted formant frequencies. We are aware that manually extracted measurements by experts are not free of errors, due to a possible "expert bias" effect [34]; hand measurements remain however the best reference, since analog spectrum analyzers are no longer generally available. Fig. 5 shows the waveform of the analyzed speech sample, while Fig. 6 contains the corresponding spectrogram computed with xkl.

Formants were measured at two different time points within the vowel, the first, T1 = 245.5 ms, corresponding to the middle time point and the second, T2 = 285.5 ms, corresponding to a time point about three quarters of the way through the region of voicing associated to the vowel (see red dotted lines on Fig. 5). As shown in the previous Section, it is straightforward to plot spectrum slices in xkl; Fig. 7 shows the DFT spectrum and the smoothed spectrum computed at t = T1, while Fig. 8 shows the same two spectra computed at t = T2. Both figures show the information returned by xkl on each estimated formant, including frequency and amplitude; the figures also indicate in red the manually estimated second formant. The manually extracted values were estimated by looking at the amplitudes of the individual harmonics throughout the region of voicing. The first formant F1 is rather stable across the entire vowel duration, as also highlighted by the spectrogram in Fig. 6. As a result, xkl estimates $F1 = 291$ Hz at T1 and $F1 = 287$ Hz at T2; these values are very close to the manually estimated value, that is, $F1 = 288$ Hz at both T1 and T2.

The estimation of the second formant is not as straightforward. At T1, xkl returns two possible candidates for F2: $F2_l = 2888$ Hz and $F2_h = 3115$ Hz (note the presence of the asterisk in the figure), in correspondence to two neighboring harmonics. On this basis F2 can be estimated by computing a weighted average of the two frequencies, with weights associated to the corresponding amplitudes, $A2_l = 67$ dB vs. $A2_h = 69$ dB, also provided by xkl. This calculation is shown in (1).

$$F2 = \frac{A2_l * F2_l + A2_h * F2_h}{A2_l + A2_h} = \frac{67 * 2888 + 69 * 3115}{67 + 69} = 3003 \text{ Hz} \tag{1}$$

This value is remarkably close to the manually estimated value $F2 = 3001$ Hz. At T2, xkl returns a single value $F2 = 3112$ Hz; this value is within 15 Hz from the manually estimated value $F2 = 3098$ Hz.

Table 1 summarizes the results of the comparison between manual estimation and xkl, and highlights that xkl provides an accurate formant estimation even in the challenging case of a young speaker with high pitch.

4. Impact

The updated availability of the xkl software offers new opportunities for carrying out the search for acoustic cues in the speech signal. As discussed above, xkl provides accurate values of formant frequencies and this capacity will facilitate our ability to sort out complex questions about e.g. the cues to rounding

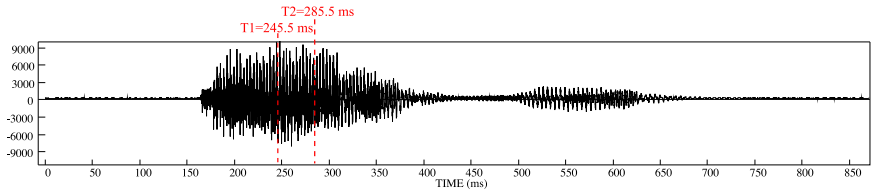

Fig. 5. Waveform for the bisyllabic sequence [izi] analyzed in Section 3.2. Time points where formants were computed are shown on figure and correspond to T1 = 245.5 ms and T2 = 285.5 ms.

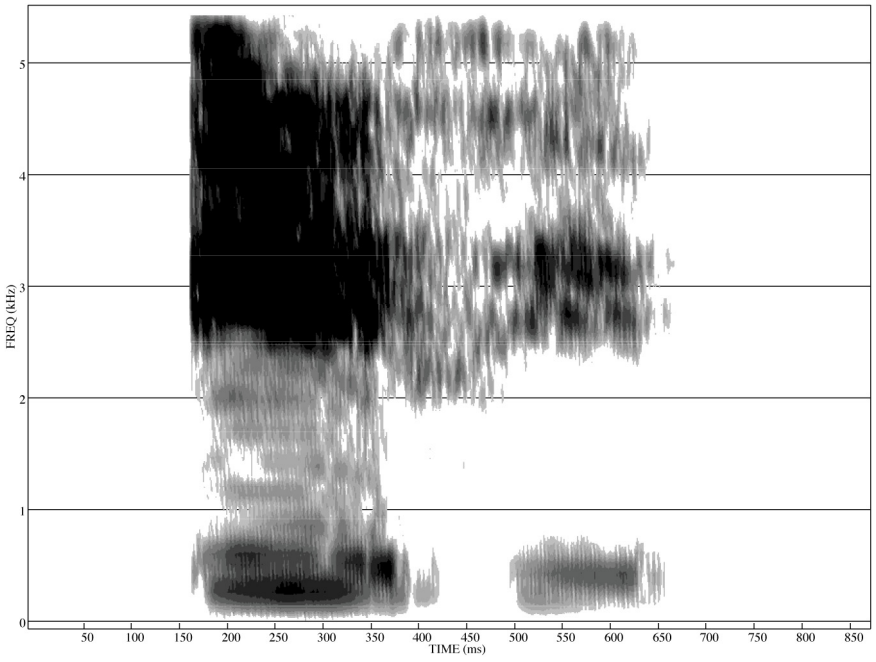

Fig. 6. Spectrogram computed with xkl for the bysillabic sequence [izi] analyzed in Section 3.2.

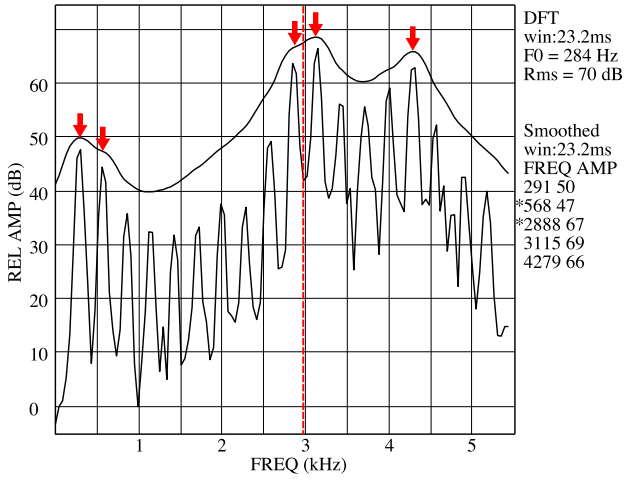

Fig. 7. Spectrum window obtained from xkl for the considered waveform at T = 245.5 ms; arrows indicate the formants estimated by xkl, with frequencies and corresponding amplitudes listed at the right of the plot, while the red dashed line indicates the manually estimated value of second formant F2 = 3004 Hz.

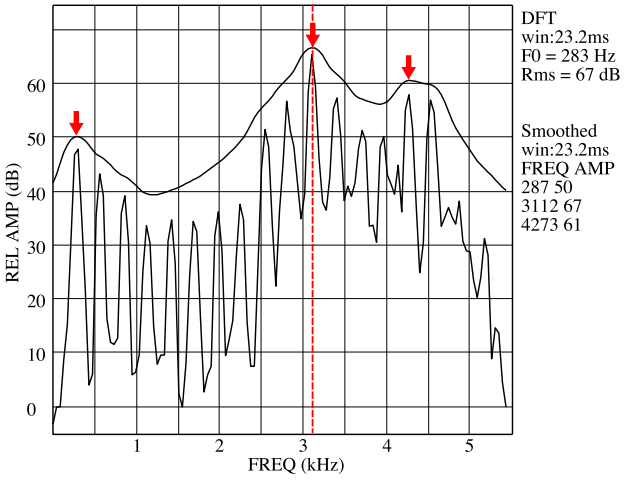

Fig. 8. Spectrum window obtained from xkl for the considered waveform at T = 285.5 ms; arrows indicate the formants estimated by xkl, with frequencies and corresponding amplitudes listed at the right of the plot, while the red dashed line indicates the manually estimated value of second formant F2 = 3098 Hz.

and nasalization, where a granularity of a few Hertz might be important [4,35]. In addition the possibility of computing the DFT with up to 4096 points, as previously mentioned, allows xkl to compute the spectrum of the signal with a frequency resolution of a few Hertz at the sampling rates typically used in recording a speech signal, opening the way for unveiling hidden acoustic manifestations.

As these examples suggest, a particularly valuable aspect of xkl is that it was specifically designed to perform acoustical analysis of phonologically relevant aspects of the speech signal in a way

Table 1
Comparison between F1 and F2 manual measurements and corresponding estimated values provided by xkl.

Time (ms)	F1 manual (Hz)	F1 xkl (Hz)	F2 manual (Hz)	F2 xkl (Hz)
245.5	288	291	3004	3003
285.5	288	287	3098	3112

that is both highly precise and also extremely convenient for the user. As a result, the user can compute spectral slices and spectrograms using just a few clicks, as well as (see previous section) measure acoustic parameters directly on the waveform or the spectrum with very fine resolution.

The resulting friendliness of the user interface and the ease in performing measurements in both the time and frequency domains makes xkl a particularly valuable tool for carrying out measurements on large scale datasets. xkl was used for example to collect measurements on a database for lexical gemination published in [36], described in [37] and investigated in [19,20]. xkl was also used for the analysis of lexical and syntactic gemination in the LaMIT database [30], described in [31] and used to investigate the phenomenon of double bursts in geminated Italian stop consonants in [21].

5. Conclusions

In this paper, a legacy speech analysis software named xkl, developed by Dennis Klatt, was presented and described in its current, revised version that can run on modern operating systems. xkl is an outstanding tool for spectral analysis; in particular, as highlighted in the example section, it provides accurate estimation of formant frequencies and of other aspects of the speech signal that are critical for understanding human speech perception and production. Our intention in writing this paper is to spread the word to the speech community regarding the availability of xkl in its modern version, under an GPL-2.0 open source license that provides free access to the source code for both commercial and non-commercial use.

The main contribution of this work is to reintroduce xkl to the speech research community, and to make its source code freely and permanently available. The current version is a first step toward the development of future releases of xkl that will proceed along two lines: first, a new Graphical User Interface based on GTK software libraries will be developed, that will make installation and use more intuitive; second, new algorithms for formant estimation will be made available, incorporating recent advances in spectrum analysis theory [38].

The above developments will be carried out also thanks to collaborative exchanges that we are confident will arise now that the software is available again.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

The authors wish to thank their institutions (Sapienza, Italy, grants RP11916B88F1A517, RP120172B3612D94, RP1221815D9D921C, and MIT, USA) for supporting the establishment of the framework that allowed the development of this collaborative project.
The work of J.-Y. Choi and S. Shattuck-Hufnagel was partially supported by the NSF, USA, with grants 1827598 (2018.9) and 1651190 (2016.9).

References

[1] Jelinek F. Continuous speech recognition by statistical methods. Proc IEEE 1976;64(4):532–56. http://dx.doi.org/10.1109/PROC.1976.10159.
[2] Deng L, Li X. Machine learning paradigms for speech recognition: An overview. IEEE Trans Audio Speech Lang Process 2013;21(5):1060–89. http://dx.doi.org/10.1109/TASL.2013.2244083.
[3] Pieraccini R. AI assistants. Cambridge, MA, USA: The MIT Press; 2021.
[4] Stevens KN. Acoustic phonetics. Cambridge, MA, USA: The MIT Press; 2000.
[5] Riches N, Loucas T, Baird G, Charman T, Simonoff E. Non-word repetition in adolescents with specific language impairment and autism plus language impairments: A qualitative analysis. J Commun Disorders 2011;44(1):23–36. http://dx.doi.org/10.1016/j.jcomdis.2010.06.003.
[6] Shriberg LD, Lohmeier HL, Campbell TF, Dollaghan CA, Green JR, Moore CA. A nonword repetition task for speakers with misarticulations: The syllable repetition task (SRT). J Speech Lang Hearing Res 2009;52(5):1189–212. http://dx.doi.org/10.1044/1092-4388(2009/08-0047).
[7] Gibbon FE, Lee A. Preface to the special issue on covert contrasts. Clin Linguist Phonetics 2017;31(1):1–3. http://dx.doi.org/10.1080/02699206.2016.1254684.
[8] Proceedings of the 11th pronunciation in second language learning and teaching conference. 2019, Avalable at https://iastate.app.box.com/v/pslltproceedings11. [Accessed on 30 June 2023].
[9] Labov W. A sociolinguistic perspective on sociophonetic research. J Phonet 2006;34(4):500–15. http://dx.doi.org/10.1016/j.wocn.2006.05.002.
[10] Boersma P, Weenink D. Praat: Doing phonetics by computer [computer program]. 2023, Version 6.3.10 (released on May 3, 2023), available at https://www.fon.hum.uva.nl/praat/. [Accessed on 30 June 2023].
[11] Wavesurfer software. 2023, Version 1.8.8.p6 (released on May 7, 2020), available at https://sourceforge.net/projects/wavesurfer/. [Accessed on June 30, 2023].
[12] VoiceSauce software. 2023, Version v1.37 (released on June 2, 2020), available at http://www.phonetics.ucla.edu/voicesauce/. [Accessed on 30 June 2023].
[13] Blumstein SE, Stevens KN. Acoustic invariance in speech production: Evidence from measurements of the spectral characteristics of stop consonants. J Acoust Soc Am 1979;66(4):1001–17. http://dx.doi.org/10.1121/1.383319.
[14] Pisoni DB, Remez RE, editors. The handbook of speech perception. Blackwell Publishing Ltd; 2005.
[15] KlSyn - C language port. 2023, Version 3.0.4 (released on May 9, 1994), available at https://www.cs.cmu.edu/afs/cs/project/ai-repository/ai/areas/speech/systems/klatt/0.html. [Accessed on 30 June 2023].
[16] KlSyn - Python 3 language port. 2023, Version released on October 7, 2021, available at https://github.com/rsprouse/klsyn. [Accessed on 30 June 2023].
[17] KlSyn - JavaScript language port. 2023, Version released on December 14, 2022, available at https://github.com/chdh/klatt-syn. [Accessed on 30 June 2023].
[18] He D, Lim BP, Yang X, Hasegawa-Johnson M, Chen D. Acoustic landmarks contain more information about the phone string than other frames for automatic speech recognition with deep neural network acoustic model. J Acoust Soc Am 2018;143(6):3207–19. http://dx.doi.org/10.1121/1.5039837.
[19] Di Benedetto M-G, De Nardis L. Consonant gemination in Italian: The nasal and liquid case. Speech Commun 2021;133:62–80. http://dx.doi.org/10.1016/j.specom.2021.07.006.
[20] Di Benedetto M-G, De Nardis L. Consonant gemination in Italian: The affricate and fricative case. Speech Commun 2021;134:86–108. http://dx.doi.org/10.1016/j.specom.2021.07.005.
[21] Di Benedetto M-G, Shattuck-Hufnagel S, De Nardis L, Budoni S, Arango J, Chan I, DeCaprio A. Lexical and syntactic gemination in Italian consonants—Does a geminate Italian consonant consist of a repeated or a strengthened consonant? J Acoust Soc Am 2021;149(5):3375–86. http://dx.doi.org/10.1121/10.0004987.
[22] Igeta T, Arai T. Dominance of lower formants of Korean vowels /o/–/u/ in perceptual identification by Seoul dialect listeners. Acoust Sci Technol 2019;40:56–8. http://dx.doi.org/10.1250/ast.40.56.

[23] Igeta T, Sonu M, Arai T. Sound change of /o/ in modern Seoul Korean: Focused on relations with acoustic characteristics and perception. Phonet Speech Sci 2014;6(3):109–19. http://dx.doi.org/10.13064/KSSS.2014.6.3.109.
[24] Tomaru K, Arai T. Discrimination of /ra/ and /la/ speech continuum by native speakers of English under nonisolated conditions. Acoust Sci Technol 2014;35(5):251–9. http://dx.doi.org/10.1250/ast.35.251.
[25] Motif libraries. 2023, Version 2.3.8 (released on December 5, 2017), available at https://sourceforge.net/projects/motif/. [Accessed on 30 June 2023].
[26] Xkl licensing page. 2023, Available at http://acts.ing.uniroma1.it/xkl.php. [Accessed on 30 June 2023].
[27] Klatt DH. Software for a cascade/parallel formant synthesizer. J Acoust Soc Am 1980;67(3):971–95. http://dx.doi.org/10.1121/1.383940.
[28] Fant G. Acoustic theory of speech production. Mouton, The Hague; 1960.
[29] Stevens KN. On the quantal nature of speech. J Phonet 1989;17(1):3–45. http://dx.doi.org/10.1016/S0095-4470(19)31520-7.
[30] Di Benedetto M-G, De Nardis L, Shattuck-Hufnagel S, Choi J-Y. The LaMIT database: A read speech corpus for acoustic studies of the Italian language toward lexical access based on the detection of landmarks and other acoustic cues to features. 2022, http://dx.doi.org/10.17632/sjwb9hymhn.2, URL https://data.mendeley.com/datasets/sjwb9hymhn/2.
[31] Di Benedetto M-G, Shattuck-Hufnagel S, Choi J-Y, De Nardis L, Arango J, Chan I, DeCaprio A, Budoni S. The LaMIT database: A read speech corpus for acoustic studies of the Italian language toward lexical access based on the detection of landmarks and other acoustic cues to features. Data In Brief 2022;42:108275. http://dx.doi.org/10.1016/j.dib.2022.108275.
[32] Di Benedetto M-G, Shattuck-Hufnagel S, Choi J-Y, De Nardis L, Arango J, Chan I, DeCaprio A. Lexical access model for Italian – modeling human speech processing: Identification of words in running speech toward lexical access based on the detection of landmarks and other acoustic cues to features. 2021, http://dx.doi.org/10.48550/ARXIV.2107.02720, arXiv URL https://arxiv.org/abs/2107.02720.
[33] Stevens KN. Toward a model for lexical access based on acoustic landmarks and distinctive features. J Acoust Soc Am 2002;111(4):1872–91. http://dx.doi.org/10.1121/1.1458026.
[34] Whalen DH, Chen W-R, Shadle CH, Fulop SA. Formants are easy to measure; Resonances, not so much: Lessons from Klatt (1986). J Acoust Soc Am 2022;152(2):933–41. http://dx.doi.org/10.1121/10.0013410.
[35] Chen MY. Nasal detection module for a knowledge-based speech recognition system. In: Sixth international conference on spoken language processing. 2000, http://dx.doi.org/10.21437/ICSLP.2000-892.
[36] Di Benedetto M-G, De Nardis L. The GEMMA speech database: VCV and VCCV words for the acoustic analysis of consonants and lexical gemination in Italian. Mendeley; 2022, http://dx.doi.org/10.17632/DM5N5DZRP2.1, URL https://data.mendeley.com/datasets/dm5n5dzrp2/1.
[37] Di Benedetto M-G, De Nardis L. The GEMMA speech database: VCV and VCCV words for the acoustic analysis of consonants and lexical gemination in Italian. Data In Brief 2022;43:108373. http://dx.doi.org/10.1016/j.dib.2022.108373.
[38] Fulop SA. Accuracy of formant measurement for synthesized vowels using the reassigned spectrogram and comparison with linear prediction. J Acoust Soc Am 2010;127(4):2114–7. http://dx.doi.org/10.1121/1.3308476.