The Journal of Open Source Software

VoxDose: An Open-Source Software for Vocal Dose Analysis

Tiago Lima Bicalho Cruz [1¶] **and Pasquale Bottalico** [2]

1 Fonotech Academy, Belo Horizonte, Brazil **2** Department of Speech and Hearing Science, University of Illinois at Urbana–Champaign, United States ¶ Corresponding author

DOI: 10.xxxxxx/draft

Software

- Review
- Repository
- Archive

Editor:

Submitted: 09 October 2025
Published: unpublished

Introduction

The human voice is a complex biological and acoustic system that enables speech, singing, and communication across professional and social domains. For occupational voice users—including teachers, singers, actors, broadcasters, and call-center operators—the voice is the primary tool of the trade, and excessive vocal load can lead to fatigue, dysphonia, and long-term injury [Titze et al. (1997); @Hunter2016; @Vilkman2004]. To address these risks, researchers have developed the concept of vocal dose, which quantifies the cumulative vibration exposure of the vocal folds in a manner analogous to radiation or chemical exposure [Titze et al. (2003)].

Vocal dose measures include: time dose (Dt), or total phonation time; cycle dose (Dc/VLI), the total number of oscillations; distance dose (Dd), the cumulative distance traveled by the vocal folds; energy dissipation dose (De), the mechanical energy absorbed by tissue; and radiated energy dose (Dr), the acoustic energy radiated to the environment [Titze et al. (2003); @Svec2004]. These parameters provide a multidimensional perspective on vocal load, going beyond simple averages of fundamental frequency (F0) and sound pressure level (SPL).

Since their formalization, vocal dose measures have been applied to a wide range of populations and contexts. Teachers have been the most frequently studied group, with evidence showing that classroom noise, reverberation, and lack of amplification significantly increase vocal doses and risk of dysphonia [Astolfi et al. (2012); @Bottalico2012; @Assad2019; @Rabelo2019]. Studies confirm that teachers typically spend around 25–27% of their working day phonating, corresponding to hundreds of thousands of vocal fold oscillations and several kilometers of cumulative tissue displacement [Astolfi et al. (2012); @Bottalico2012]. Amplification has been shown to significantly reduce cycle and distance doses in dysphonic teachers [Assad et al. (2019)], while classroom noise increases fundamental frequency, vocal intensity, percentage of phonation (time dose), and cycle dose through the Lombard effect [Rabelo et al. (2019)].

Vocal dose analysis has also expanded into singing and performance [Carroll et al. (2006)]. Professional singers display higher doses than untrained controls, but with more efficient respiratory–phonatory coordination [Cunsolo et al. (2022)]. In contemporary musical theatre, singers accumulate extremely high cycle and distance doses, particularly women, due to stylistic demands of chest-dominant phonation. Self-perception, as measured by the Evaluation of the Ability to Sing Easily (EASE) questionnaire, showed poor correlation with objective exposure [Zuim et al. (2024)].

Technological advances have shaped the field. The NCVS dosimeter (Titze & Hunter, 2004) set the foundation for ambulatory monitoring using neck-surface accelerometers. Commercial systems (APM, VoxLog, VocaLog, Voice-Care) have since become available, though cost and limited parameter extraction remain barriers [Hunter (2016)]. Recent innovations include smartphone-based systems [Castellana et al. (2018); @Hunter2016; @Mehta2015] and low-cost DIY solutions [Bottalico & Nudelman (2023)], making vocal dosimetry more accessible. Research has also integrated subglottal impedance-based inverse filtering (IBIF) to estimate aerodynamic parameters [Mehta et al. (2015)] and accelerometer-based prediction of subglottal pressure [Fryd et al. (2016)], expanding the scope of ambulatory monitoring.

Systematic reviews highlight consistent findings: high vocal doses are linked to teaching, noisy environments, dysphonia, and vocal fatigue, while amplification and vocal rest reduce exposure [Assad et al. (2017)]. Collectively, these studies demonstrate that vocal dose measures are powerful tools for understanding occupational voice use, guiding preventive strategies, and supporting clinical decision-making.

Statement of Need

Despite two decades of research into vocal dose, existing tools for analysis remain either costly, closed-source, or inaccessible to clinicians and researchers outside specialized laboratories. Teachers, clinicians, and voice scientists often rely on proprietary dosimeters or complex workflows requiring Praat scripting, MATLAB, or non-standardized pipelines. There is a pressing need for an open-source, user-friendly, and reproducible software that implements validated vocal dose metrics, integrates SPL and F0 calibration, and exports standardized results for research, pedagogy, and clinical practice.

VoxDose addresses this gap by providing a free, open-source application that computes all major vocal dose measures (Dt, VLI, Dd, De, Dr) from recorded audio, with calibration features, visualization, and batch analysis. By bridging the gap between advanced voice science and practical applications, VoxDose democratizes access to vocal dosimetry for researchers, clinicians, and educators worldwide.

Installation

VoxDose is distributed as open-source Python code. The recommended installation procedure is as follows:

Prerequisites

- Python 3.9 or later (tested on Windows, macOS, Linux)
- Required libraries: numpy, scipy, matplotlib, pandas, PySide6, praat-parselmouth, openpyxl

Step-by-step

1. Clone the repository:
2. git clone https://github.com/tiagolbc/voxdose.git
3. cd voxdose
4. Install dependencies:
5. pip install -r requirements.txt
6. Run the application:
7. python main_gui.py

Software Description

Purpose and Features

- Frame-by-frame analysis of Sound Pressure Level (SPL, dBA) and fundamental frequency (F0, Hz) from WAV or MP3 recordings.
- Calculation of vocal dose metrics:
  - Dt (phonation time),
  - VLI (Vocal Loading Index, cycles $\times$ 1000),

The Journal of Open Source Software

- Dd (distance dose, meters),
- De (energy dissipation dose, joules),
- Dr (radiated energy dose, joules),
- plus normalized measures and SPL/F0 statistics.

- Calibration (30 cm vowel + SLM): Users record a sustained vowel at 30 cm while reading SPL on a calibrated sound level meter at the same position. VoxDose takes the calibration audio file and the measured SPL (dBA), computes a calibration constant, and applies it to the analysis. During processing, users can report SPL at 30 cm (default) or re-reference to 50 cm:

$$SPL_{target} = SPL_{measured} - 20\log_{10}\left(\frac{d_{cal}}{d_{target}}\right)$$

```
-   This correction ensures that all SPL values - and consequently all derived dose meas
    (Dt, VLI, Dd, De, Dr) - are expressed in absolute, physically valid units consistent
```

- Interactive Graphical User Interface (GUI) with:
    - File selection (voice recordings and calibration recording for SPL).
    - Input fields for SPL calibration, microphone distance, and F0 search range.
    - Sex-specific analysis paths (male, female, or "other") affecting biomechanical scaling.
    - Export options for results in Excel (frame-by-frame and summary doses).
- Visualization: SPL and F0 time series with automatic mean annotation; summary plots saved alongside results; spectrogram and pitch tracking available for advanced inspection.

Implementation and Architecture

Core modules

- **dosi.py** – Implements all vocal dose equations (Dt, VLI, Dd, De, Dr), following Bottalico's MATLAB framework with sex-specific physiological scaling.
- **spl_fast.py / spl_fast_c_th.py** – Frame-by-frame SPL computation kernels, with or without calibration constant.
- **stima_livello.py** – FFT-based SPL estimator used in calibration and validation.
- **sp_pitch_praat.py and sp_pitch_track_praat.py** – Parselmouth wrappers for Praat autocorrelation-based pitch extraction, used for both sustained vowels and connected speech.
- **sp_cpps.py** – CPPS estimation for connected speech, computed every 5 s on voiced segments after pause removal.
- **analyze_wav_spl_f0.py** – Central analysis routine: integrates SPL and F0 pipelines, runs calibration, synchronizes arrays, removes silences, and exports both frame-level data and summary dose results.
- **main_gui.py** – PySide6-based graphical interface providing file selection, calibration entry, analysis controls, plotting, and export.
- **splash.py** – Startup splash screen with license and credits.

Analysis workflow

1. **Input**: user selects voice file(s), optional calibration file, enters measured SPL, and sets distance options.
2. **Preprocessing**: SPL is computed with the calibration constant; F0 is tracked with Praat autocorrelation; silences (<50 dBA) are masked.
3. **Synchronization**: SPL and F0 arrays are aligned at 50 ms frame resolution.
4. **Dose computation**: dosi.py integrates the frame-based values into cumulative dose metrics.
5. **Export**: results are written into two Excel files per recording:
   - [basename].xlsx (time, SPL, F0 per frame).
   - [basename]_VocalDoses.xlsx (summary table with all dose metrics).
6. **Visualization**: SPL and F0 plots with mean annotations, plus summary PNG plots of the dose metrics.

Architecture

The modular structure separates signal processing (pitch, SPL, cepstrum), mathematical modeling (dose equations), and user interface (GUI, splash, exports). This design allows VoxDose to be easily extended with new acoustic features (e.g., HNR, alpha ratio) or alternative pitch/SPL methods, while maintaining a clean and reproducible analysis pipeline.

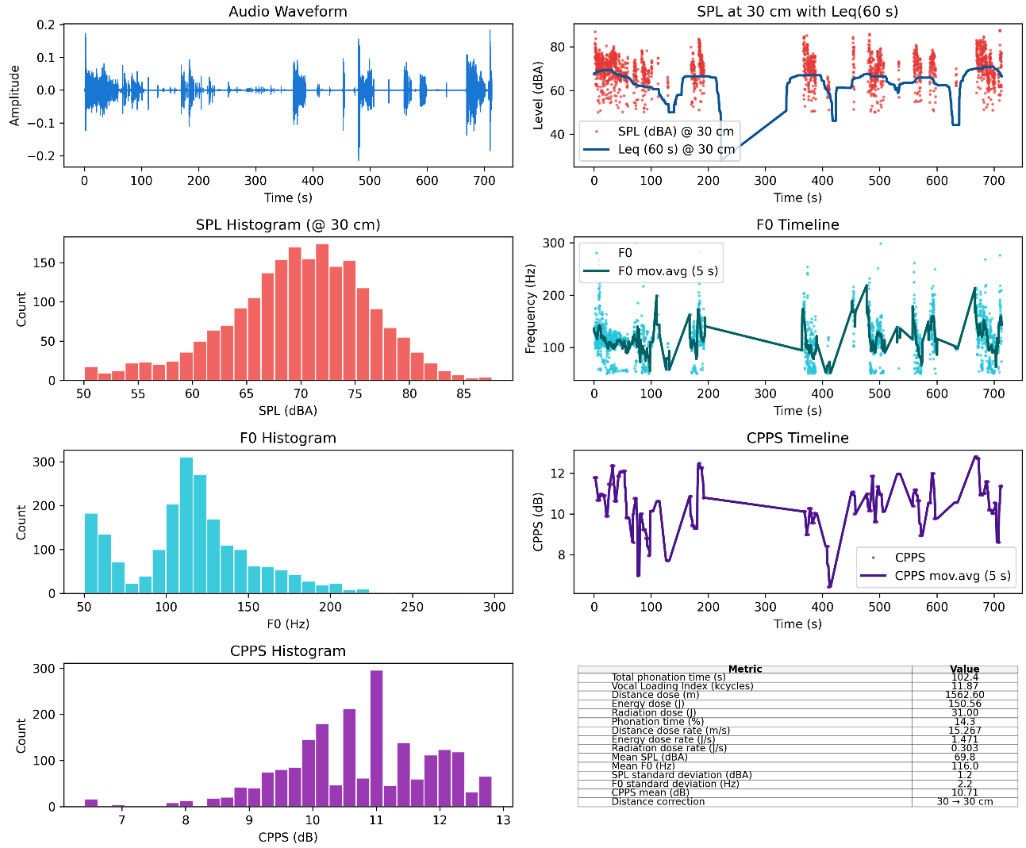
Illustrative Examples

To illustrate the functionality of VoxDose, we present one analysis example recorded by the first author. The test consisted of a sustained vowel calibration followed by a short connected-speech passage.

The first step is shown in Figure 1, where the GUI allows the user to select the voice recording, load the calibration file, enter the SPL value measured with the sound level meter, and choose whether results should be reported at 30 cm or re-referenced to 50 cm. The GUI provides a simple and intuitive workflow, designed for researchers, clinicians, and educators who may not have programming experience.

Figure 1. VoxDose graphical interface showing file selection, calibration entry, distance options, and analysis controls.

Once the calibration and analysis are executed, VoxDose produces both frame-level and summary outputs. Figure 2 illustrates the visualization of SPL and F0 time series, with mean values automatically annotated, as well as the export of cumulative vocal dose measures (Dt, VLI, Dd, De, Dr) to an Excel file. This integration of graphical and tabular results provides a comprehensive view of vocal load, making it possible to interpret individual phonatory behavior in absolute units.



Metric	Value
Total phonation time (s)	102.4
Vocal Loading Index (kcycles)	11.87
Distance dose (m)	1562.60
Energy dose (J)	150.56
Radiation dose (J)	31.00
Phonation time (%)	14.3
Distance dose rate (m/s)	15.267
Energy dose rate (J/s)	1.471
Radiation dose rate (J/s)	0.303
Mean SPL (dBA)	69.8
Mean F0 (Hz)	116.0
SPL standard deviation (dBA)	1.2
F0 standard deviation (Hz)	2.2
CPPS mean (dB)	10.71
Distance correction	30 → 30 cm

Figure 2. Example output generated by VoxDose: SPL and F0 curves over time with mean annotations, alongside a summary table of vocal dose metrics.

Comparison with Existing Tools

Vocal dosimetry research has historically relied on hardware-based systems. Early solutions, such as speech timers and voice accumulators [Ryu et al. (1983); @Rantala1999], initially focused on quantifying phonation time, with later advancements incorporating measures like the total number of vocal fold oscillations. More advanced devices, like the KayPENTAX Ambulatory Phonation Monitor (APM 3200) and the NCVS dosimeter [Švec et al. (2004)], introduced accelerometer-based monitoring of F0, SPL, and phonation time, enabling the first large-scale studies of vocal load. Commercial successors—including VocaLog2 (Griffin Labs), VoxLog (Sonvox AB), and Voice-Care (PR.O.VOICE)—continue this tradition, but remain hardware products with proprietary data formats, high costs, and limited analytical flexibility [Hunter (2016)].

Recent developments have explored DIY dosimeters [Bottalico & Nudelman (2023)] and smartphone-based solutions [Mehta et al. (2015)], expanding accessibility but still tied to hardware for data collection. These devices generate raw data streams (accelerometer, SPL, F0) that must then be processed into meaningful measures of vocal dose.

VoxDose does not compete with these hardware systems, but instead provides a software-only solution for analysis and visualization of vocal dose metrics. It is designed to:

- Process pre-recorded audio files (e.g., WAV) rather than raw accelerometer signals.
- Implement validated mathematical models of vocal dose (Dt, VLI, Dd, De, Dr) [Titze et al. (2003); @Bottalico2012].
- Integrate SPL calibration against sound level meter values, ensuring external consistency.

- Offer open-source accessibility, in contrast to closed proprietary ecosystems.
- Facilitate research and clinical practice where hardware dosimeters are not available, or where recordings need retrospective analysis.

In this sense, VoxDose is best understood as a complementary tool: hardware dosimeters capture real-life ambulatory data, while VoxDose provides a transparent, reproducible environment to analyze, visualize, and export vocal dose measures from recorded speech or singing tasks.

Acknowledgements

The mathematical framework for vocal dose measures was originally developed in MATLAB by Prof. Pasquale Bottalico (University of Illinois at Urbana-Champaign). His algorithms served as the foundation for the present Python reimplementation in VoxDose, ensuring continuity and reproducibility.

VoxDose is part of the FonoTech Academy open-source ecosystem, dedicated to making advanced voice science tools accessible, transparent, and reproducible for researchers, clinicians, and educators worldwide.

Assad, J. P., Gama, A. C. C., Santos, J. N., & Magalhães, M. de C. (2019). The effects of amplification on vocal dose in teachers with dysphonia. *Journal of Voice*, *33*(1), 73–79. https://doi.org/10.1016/j.jvoice.2017.09.011

Assad, J. P., Magalhães, M. de C., Santos, J. N., & Gama, A. C. C. (2017). Vocal dose: An integrative literature review. *Revista CEFAC*, *19*(3), 429–438. https://doi.org/10.1590/1982-021620171932617

Astolfi, A., Bottalico, P., Accornero, A., Garzaro, M., Nadalin, J., & Giordano, C. (2012, June). Relationship between vocal doses and voice disorders on primary school teachers. *Proceedings of Euronoise 2012*.

Bottalico, P., & Nudelman, C. J. (2023). Do-it-yourself voice dosimeter device: A tutorial and performance results. *Journal of Speech, Language, and Hearing Research*. https://doi.org/10.1044/2023_JSLHR-23-00060

Carroll, T. L., Nix, J., Hunter, E., Emerich, K., Titze, I., & Abaza, M. (2006). Objective measurement of vocal fatigue in classical singers: A vocal dosimetry pilot study. *Otolaryngology–Head and Neck Surgery*, *135*(4), 595–602. https://doi.org/10.1016/j.otohns.2006.05.003

Castellana, A., Carullo, A., Corbellini, S., & Astolfi, A. (2018). Discriminating pathological voice from healthy voice using cepstral peak prominence smoothed distribution in sustained vowel. *IEEE Transactions on Instrumentation and Measurement*, *67*(1), 214–222. https://doi.org/10.1109/TIM.2017.2781958

Cunsolo, F., Ottaviani, V., Capobianco, S., Calcinoni, O., & Dellacà, R. L. (2022). Simultaneous monitoring of vocal doses and breathing patterns in professional singers. *Computers in Biology and Medicine*, *144*, 105352. https://doi.org/10.1016/j.compbiomed.2022.105352

Fryd, A. S., Van Stan, J. H., Hillman, R. E., & Mehta, D. D. (2016). Estimating subglottal pressure from neck-surface acceleration during normal voice production. *Journal of Speech, Language, and Hearing Research*, *59*, 1335–1345. https://doi.org/10.1044/2016_JSLHR-S-15-0430

Hunter, E. J. (2016, September). Vocal dose measures: General rationale, traditional methods and recent advances. *I International Conference on Occupational Voice*.

Mehta, D. D., Van Stan, J. H., Zañartu, M., Ghassemi, M., Guttag, J. V., Espinoza, V. M., Cortés, J. P., Cheyne, H. A. I., & Hillman, R. E. (2015). Using ambulatory voice monitoring to investigate common voice disorders: Research update. *Frontiers in Bioengineering and Biotechnology*, *3*, 155. https://doi.org/10.3389/fbioe.2015.00155

Rabelo, A. T. V., Santos, J. N., Souza, B. O., Gama, A. C. C., & Magalhães, M. de C. (2019). The influence of noise on the vocal dose in women. *Journal of Voice*, *33*(2), 214–219. https://doi.org/10.1016/j.jvoice.2017.10.025

Ryu, S., Komiyama, S., Kannae, S., & Watanabe, H. (1983). A newly devised speech accumulator. *Journal for Oto-Rhino-Laryngology and Its Related Specialties*, *45*(2), 108–114. https://doi.org/10.1159/000275632

Švec, J. G., Hunter, E. J., Popolo, P. S., Rogge-Miller, K., & Titze, I. R. (2004). *The calibration and setup of the NCVS dosimeter* (Online Technical Memo No. 11). National Center for Voice; Speech.

Titze, I. R., Lemke, J., & Montequin, D. (1997). Populations in the u.s. Workforce who rely on voice as a primary tool of trade: A preliminary report. *Journal of Voice*, *11*(3), 254–259. https://doi.org/10.1016/S0892-1997(97)80002-1

Titze, I. R., Švec, J. G., & Popolo, P. S. (2003). Vocal dose measures: Quantifying accumulated vibration exposure in vocal fold tissues. *Journal of Speech, Language, and Hearing Research*, *46*(4), 919–932. https://doi.org/10.1044/1092-4388(2003/072)

Zuim, A. F., Stewart, C. F., & Titze, I. R. (2024). Vocal dose and vocal demands in contemporary musical theatre. *Journal of Voice*, *38*(1), 53–68. https://doi.org/10.1016/j.jvoice.2021.08.006