SoftwareX 7 (2018) 74–80

Contents lists available at ScienceDirect

SoftwareX

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

Original software publication

FonaDyn — A system for real-time analysis of the electroglottogram, over the voice range

Sten Ternström *, Dennis Johansson, Andreas Selamtzis

Department of Speech, Music and Hearing, School of Electrical Engineering and Computer Science, KTH Royal Institute of Technology, SE-100 44, Stockholm, Sweden

ARTICLE INFO

Article history:
Received 24 October 2017
Received in revised form 19 January 2018
Accepted 8 March 2018

Keywords:
Electroglottography
Voice analysis
Clustering
Sample entropy
Voice range profile
SuperCollider

ABSTRACT

From soft to loud and low to high, the mechanisms of human voice have many degrees of freedom, making it difficult to assess phonation from the acoustic signal alone. FonaDyn is a research tool that combines acoustics with electroglottography (EGG). It characterizes and visualizes in real time the dynamics of EGG waveforms, using statistical clustering of the cycle-synchronous EGG Fourier components, and their sample entropy. The prevalence and stability of different EGG waveshapes are mapped as colored regions into a so-called voice range profile, without needing pre-defined thresholds or categories. With appropriately 'trained' clusters, FonaDyn can classify and map voice regimes. This is of potential scientific, clinical and pedagogical interest.

Code metadata

Current code version	1.5.0
Permanent link to code/repository used of this code version	FonaDyn: Uploaded ZIP archive "FonaDynInstall-1-5-0.zip" *SuperCollider:* https://github.com/ElsevierSoftwareX/SOFTX-D-17-00078
Legal Code License	FonaDyn: EUPL 1.2, http://eupl.eu
Code versioning system used	Git
Software code languages, tools, and services used	The FonaDyn source code is cross-platform. The SuperCollider 3.8.0 interpreter is needed for running it. C++ is needed (only exceptionally) for modifying the pre-compiled FonaDyn plugins. Source code for the latter is available on request from the first author at stern@kth.se.
Compilation requirements, operating environments & dependencies	On Linux, SuperCollider itself needs to compiled. On Windows and MacOS, executable installers are available (Software metadata table). On Windows: Microsoft Visual C++ 2015 Redistributable (x86)
If available Link to developer documentation/manual	Bundled with the distribution ZIP archive
Support email for questions	stern@kth.se

Software metadata

Current software version	1.5.0
Permanent link to executables of this version	The uploaded ZIP archive includes the executable platform-specific plug-in DLL's for FonaDyn: "FonaDynInstall-1-5-0.zip"
Legal Software License	FonaDyn: EUPL 1.2 (European Union Public License v1.2)
Computing platforms/Operating Systems	Microsoft Windows 7 or higher, OS X, Linux (Linux users will have to make also a build of SuperCollider itself). The current Windows version of FonaDyn is 32-bit; it runs also on 64-bit Windows. The custom FonaDyn DSP plug-ins are written in portable C++, and pre-compiled to platform-specific DLL's. The FonaDyn install script selects the correct DLL's for the platform.
Installation requirements & dependencies	SuperCollider 3.6.6 to 3.8.1, with the SC3-plugins, at http://supercollider.github.io/download. On Windows, choose the 32-bit version of both. On Windows: Microsoft Visual C++ 2015 Redistributable (x86). Please note: SuperCollider 3.9.0 was released in January 2018. It changes the plug-in executable format, and so FonaDyn 1.5.x is not compatible with SC 3.9.x. Future versions 1.6.x and higher of FonaDyn will be compatible only with SC 3.9.x and higher, 64-bit.
If available, link to user manual–if formally published include a reference to the publication in the reference list	The user handbook is included in the distribution ZIP archive.
Support email for questions	stern@kth.se

* Corresponding author.
E-mail addresses: stern@kth.se (S. Ternström), denjoh@kth.se (D. Johansson), selamt@kth.se (A. Selamtzis).

https://doi.org/10.1016/j.softx.2018.03.002

1. Motivation and significance

1.1. Background

Human voice has very many degrees of freedom, making it difficult to quantify and assess vocal status and behavior from the acoustic signal alone. For example, the vocal folds vibrate and collide in several different ways over the soft-loud and low-high ranges of the voice. While the electroglottographic (EGG) signal [1] is familiar to many vocologists, the dynamic variation of its features has not been researched in any great detail. A system that can classify EGG pulse shapes and map them across the voice range can be expected to give important insights into normal and pathological voice in speech and in singing. Doing it in real time is key in clinical and pedagogical settings, for quick and intuitive assessment or for biofeedback. This article describes a new real-time EGG analysis system that is based entirely on open-source free software and readily available medium-cost hardware. The system is called FonaDyn (for 'phonatory dynamics').

1.2. Importance and scope

Ideally, we would like to be able to see directly the vocal folds vibrating, in 3D. The EGG signal, however, is a time-varying scalar; it represents only the instantaneous, uncalibrated contact area between the two vocal folds (VF). Understanding the VF using only the EGG is like trying to infer the changing shape and motion of a 3D object by observing only the area of its silhouette. Still, the range of plausible VF positions and shapes is rather well known, constraining the possible interpretations [2,3]. Compared to optical imaging by endoscopy, acquiring an EGG signal from a subject is non-invasive and relatively inexpensive, and the EGG offers a more direct representation of vocal fold (VF) vibration than does the acoustic microphone signal. For example, when normalized to the period time, a narrow EGG pulse indicates brief VF contact during the phonatory cycle; while wider pulses indicate firm or pressed VF contact. Such observations can be of great clinical and/or pedagogical value.

1.3. Contribution

The FonaDyn system provides the voice researcher with *maps* of various modes of phonation (reflecting how the voice is used) over all levels of vocal loudness and pitch, in a relatively short time. It does so by applying automatic real-time clustering to EGG pulse shapes, plotting the resulting clusters over a so-called voice range profile (VRP) [4]. The display is optionally complemented with the sample entropy, as a metric of phonatory instability. FonaDyn has been pre-researched [5–8] and used in pilot studies [9]. For robust deployment in the clinic or singing studio, the system will in most cases need to be customized further.

1.4. Experimental setting and usage

Readers who wish to try FonaDyn are recommended first to read chapter one of the accompanying handbook, where all aspects of setting up are described. Two software packages need to be installed: SuperCollider and FonaDyn.

Fig. 1 shows a typical setup. Normally, signal acquisition is done first: FonaDyn records the microphone and EGG signals into 2-channel WAV files (44.1 kHz, 16 bit). The recordings may be trimmed using any WAV editor, so as to remove irrelevant vocalizations or long silences. The trimmed files can then be passed back through FonaDyn for analysis. Live input and disk file input are however equivalent: FonaDyn runs its full display in either case, offering visual feedback, and the display can run live without making a recording. The analyst chooses a number of waveshape clusters that is appropriate for the range of phonatory patterns that are manifest or expected in the EGG signal. This number is necessarily found by manual experimentation, informed by the research question.

1.5. Related work

Marasek [6,10] proposed a 'rough rules' paradigm for analyzing the EGG, and implemented an off-line procedure that successfully classified modal, breathy, creaky and Parkinson voice. One major source of variation in the voice is the vocal 'registers', or 'phonatory mechanisms' [11]. The two most often discussed mechanisms are modal/chest/M1 voice, as contrasted with falsetto/head/M2 voice, which are of particular interest in singing, but also in clinical settings. Selamtzis and Ternström [6] explored the EGG for discriminating automatically between the M1 and M2 modes of phonation, and found that this is possible to a useful level of confidence, if three rather than two clusters are used. They used an offline procedure in MATLAB® [12], which has informed the design of the present system. Other examples of phonatory dimensions that can be studied with FonaDyn include vocal loudness [7] and degree of vocal fold adduction [9]. Although FonaDyn was designed with the EGG signal in mind, it can be used also with any other periodic signal derived from phonation, such as the photoglottogram (for glottal area) or a signal from a neck-mounted accelerometer.

Other efforts to characterize the EGG signal have focused mainly on deriving various scalar metrics for the pulse shape. This usually necessitates the definition of time-domain thresholds, which can be problematic [13]. One of few studies to account for the shape of the whole EGG pulse did so using principal component analysis [14]. Herbst et al. proposed the wavegram, a rich and compact visualization of the EGG waveshape over time [15], which however defers the classification task to a human observer. FonaDyn instead classifies the EGG automatically, and maps the clusters on the f_0/SPL plane.

2. Software description

2.1. Software architecture

FonaDyn is implemented in SuperCollider (SC) [16], a programming environment for performing real-time sound analysis and synthesis. Originally developed by James McCartney, it is now maintained by a lively computer music community. Its users include also scientists, who have contributed a wealth of class libraries and plug-ins, making SuperCollider seriously useful also in an audio/music/acoustics research environment.

SuperCollider is open-source freeware that is supported on Windows, Mac OSX and Linux. It has three major components: (1) a signal processing server, SCSYNTH, (2) an interpreted, object-oriented programming language, SCLANG, acting as a client of the server, and (3) an integrated development environment, SCIDE, with an editor, control windows and help system. Code for the server and for user interaction and display are all written in SCLANG, which is somewhat similar to Smalltalk, with idioms of other languages mixed in. SCLANG is profoundly object-oriented, and has many elegant, compact constructs for creative manipulation of arrays and collections, as applied in music composition. The client and server communicate internally using Open Sound Control (OSC) messages, via a network protocol. For performance, FonaDyn requires that the server and client be running on the same computer, but SuperCollider itself does not.

FonaDyn is not a stand-alone program, but a collection of classes and plug-ins that extend SCLANG. The classes are provided in source code, the plug-ins as executable DLLs. Hence FonaDyn is