Interspeech 2024
1-5 September 2024, Kos, Greece

A demonstrator for articulation-based command word recognition

João Vítor Possamai de Menezes, Arne-Lukas Fietkau, Tom Diener, Steffen Kürbis, Peter Birkholz

Institute of Acoustics and Speech Communication, Technische Universität Dresden, Germany
{joao_vitor.possamai_de_menezes, arne-lukas.fietkau, peter.birkholz}@tu-dresden.de

Abstract

This paper proposes a demonstration system capable of recognizing command words with a custom optopalatographic device that measures speech articulation. Beyond the measuring device, the proposed system is composed of two interacting computer programs - one for recording and one for the actual word recognition. Operating offline, the recording software can be used to record large datasets, which in turn are used by the recognition software to train and validate classification models. Operating online, the recording software can be used to record single words, which are classified using pre-trained classification models. The system was developed modularly, enabling different devices for measuring speech articulation and different classification models to be easily integrated. The current classification system employs dynamic time warping in combination with a pattern matching algorithm and achieves 98.38% accuracy on a single speaker 40-word corpus.

Index Terms: optopalatography, silent speech interfaces, command word recognition

1. Introduction

The acquisition and visualisation of articulatory data are essential not only for research in experimental phonetics, but also for Silent Speech Interfaces (SSI), which have gained significant research interest lately due to its promising applications, e.g., speech restoration [1]. Motivated by these applications, a variety of software capable of handling articulatory data have been developed for different sensing modalities, e.g., ultrasound [2], optopalatrography (OPG) [3] and electromagnetic articulography (EMA) [4].

The system proposed in this paper is based on the open-source software Articulatory Data Recorder (ADR) [3]. We developed the software further, resulting in the new version Articulatory Data Recorded 2.0 (ADR2), and integrated it with a custom software for command word recognition and with a custom OPG device to measure speech articulation. Compared to other articulatory sensing modalities, OPG is very robust against session and speaker variability [5], it provides direct measurements suitable for speech therapy [6] and is under continuous development [7]. To measure the tongue and lip movemement a pseudopalate assembled with lasers and optical sensors is placed inside the speakers mouth.

The goal of this paper is to propose a system capable of recording and visualizing articulatory data, and of training command word recognition models and perform inference on unseen live recorded data. The system was developed modularly with the aim of being agnostic to the type of articulatory data and to the command word recognition model.

2. Proposed system

Figure 1 shows the dataflow within the proposed system. Each of its elements are described in the next sections.

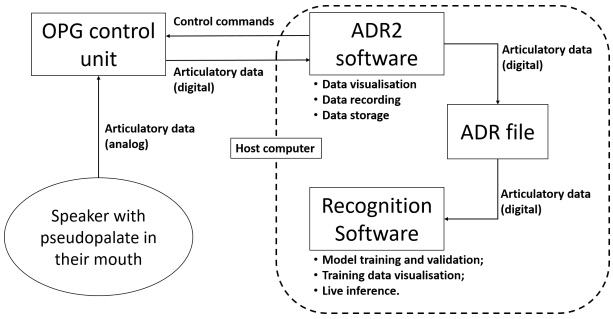

Figure 1: *Dataflow of the proposed system.*

2.1. Optopalatographic device

The OPG system used is an improved version of OPG Model 6 presented in [7]. This version uses a VCSEL and a phototransistor as an optical sensor pair. Fifteen sensor pairs are placed on a flexible printed circuit board to detect lip (2 pairs) and tongue (13 pairs) movements. As a recording device a portable control unit was developed which allows an articulation sampling rate of 100 Hz.

The resulting pseudopalate and the control unit are shown in Figure 2, as well the USB interfaces used to connect the OPG device to a computer.

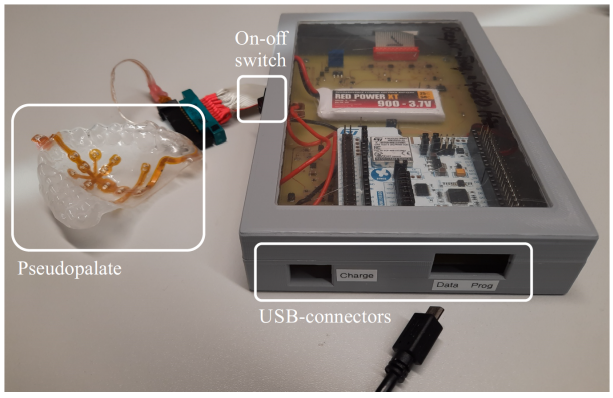

Figure 2: *OPG2023 device with attached pseudopalate.*

2.2. Recording software - Articulatory Data Recorder 2.0

The Articulatory Data Recorder 2.0 (ADR2) is capable to visualise, record and store articulatory data. It was written in C++ and uses the libraries wxWidgets 3.1.3 (`https://www.wxwidgets.org/`) for the graphical user interface (GUI), and PortAudio (`https://www.portaudio.com/`) for audio recording and playback. One of its main features is the synchronous recording of articulatory and audio data, which are generally recorded with different devices and sampling rates. The synchronization mechanism is based on two loud beeps emitted by the OPG control unit during the first and last recorded OPG frames. The two beeps are automatically detected in the audio signal (recorded by the host PC) after recording and used to time-align both signals.

Figure 3 shows the GUI of ADR2, namely the menu, the toolbar, the main area where the signals are shown, and a prompt area.

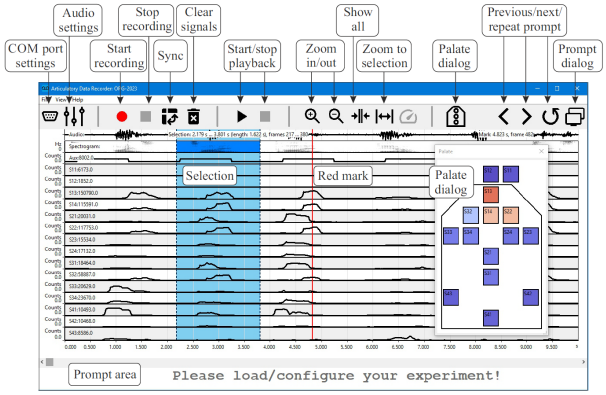

Figure 3: *Graphical User Interface of the custom recording software Articulatory Data Recorder 2.0.*

2.3. Command word recognition software

The command word recognition software enables to load and visualize articulatory data so as to train and deploy classification models. The program was developed using Python in order to integrate with machine learning libraries. It provides a GUI based on the module PySide6 (`https://pypi.org/project/PySide6/`).

Figure 4 shows the GUI of the command word recognition software, which consists of a toolbar, a settings menu, a display menu, a message box with the classification results and a status bar. In the settings menu there are options for data loading, model training and validation, and recording classification. Here is where the data recorded with Articulatory Data Recorder 2.0 are collected, as the recording path is a mandatory input. In the display menu, the words composing the loaded corpus are shown in a grid. After the classification of a recording, the words are color coded according to the results from darkest (lowest probability) to brightest (highest probability), and the word with the highest probability appears on the classification results message box.

The classification algorithm implemented so far is a Dynamic Time Warping (DTW)-based pattern matching algorithm, which achieved 98.38% accuracy for a single speaker on a 40-word German Word Corpus when assessed via 5-fold cross validation.

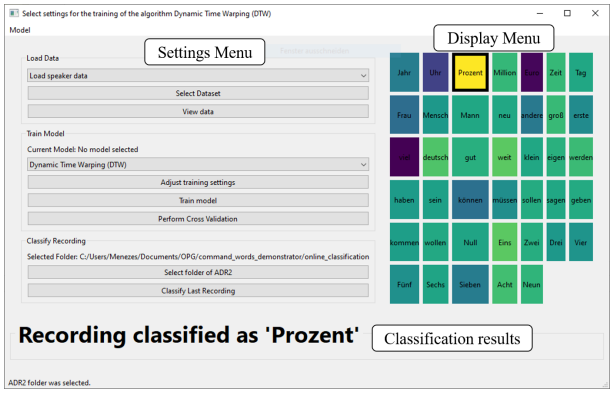

Figure 4: *Graphical User Interface of the custom command word recognition software.*

3. Discussion and outlook

The presented system successfully integrates articulatory data recording and classification based on two dedicated custom softwares and a custom OPG hardware. The modularity of the classification system allows not only the integration of various corpora and classification algorithms, but also of different speakers, who can easily record new training data with their own OPG pseudopalate and be able to operate the demonstrator.

4. Acknowledgements

The authors acknowledge the financial support by the Federal Ministry of Education and Research of Germany in the programme of “Souverän. Digital. Vernetzt.”, joint project 6G-life (project identification number: 16KISK001K), and by the European Regional Development Fund (EFRE) and the Free State of Saxony (ref. 100686372).

5. References

[1] J. A. Gonzalez-Lopez, A. Gomez-Alanis, J. M. Martín Doñas, J. L. Pérez-Córdoba, and A. M. Gomez, “Silent speech interfaces for speech restoration: A review,” *IEEE Access*, vol. 8, pp. 177 995–178 021, 2020.

[2] A. Wrench, *Articulate Assistant Advanced User Guide: Version 2.17*. Articulate Instruments Ltd, 2017.

[3] A. Wilbrandt, S. Stone, and P. Birkholz, “Articulatory Data Recorder: A Framework for Real-Time Articulatory Data Recording,” in *Proc. Interspeech 2021*, 2021, pp. 3313–3314.

[4] P. Buech, S. Roessig, L. Pagel, D. Muecke, and A. Hermes, “ema2wav: doing articulation by Praat,” in *Proc. Interspeech 2022*, 2022, pp. 1352–1356.

[5] S. Stone and P. Birkholz, “Cross-speaker silent-speech command word recognition using electro-optical stomatography,” in *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2020, pp. 7849–7853.

[6] C. Wagner, L. Stappenbeck, H. Wenzel, P. Steiner, B. Lehnert, and P. Birkholz, “Evaluation of a non-personalized optopalatographic device for prospective use in functional post-stroke dysphagia therapy,” *IEEE transactions on bio-medical engineering*, vol. 69, no. 1, 2022.

[7] P. Birkholz, S. Stone, C. Wagner, S. Kürbis, A. Wilbrandt, and M. Bosshammer, “A review of palatographic measurement devices developed at the TU Dresden from 2011 to 2022,” in *Proceedings of the 20th International Congress of Phonetic Sciences*, R. Skarnitzl and J. Volín, Eds., 2023, pp. 883–887.