THROAT MICROPHONES FOR VOCAL MUSIC ANALYSIS

Frank Scherbaum[1] **Sebastian Rosenzweig**[2] **Meinard Müller**[2] **Daniel Vollmer**[1]
Nana Mzhavanadze[3]
[1] Institut für Erd- und Umweltwissenschaften, Universität Potsdam, Germany
[2] International Audio Laboratories Erlangen, Germany
[3] Ilia State University, Tbilisi, Georgia
{fs, daniel}@geo.uni-potsdam.de,
{sebastian.rosenzweig, meinard.mueller}@audiolabs-erlangen.de

ABSTRACT

Due to the complex nature of the human voice, the computational analysis of polyphonic vocal music recordings constitutes a challenging scenario. Development and evaluation of automated music processing methods often rely on multitrack recordings comprising one or several tracks per voice. However, recording singers separately is neither always possible, nor is it generally desirable. As a consequence, producing clean recordings of individual voices for computational analysis is problematic. In this context, one may use throat microphones which capture the vibrations of a singers' throat, thus being robust to other surrounding acoustic sources. In this contribution, we sketch the potential of such microphones for music information retrieval tasks such as melody extraction. Furthermore, we report on first experiments conducted in the course of a recent project on computational ethnomusicology, where we use throat microphones to analyze traditional three-voice Georgian vocal music.

1. THROAT MICROPHONES

Uniqueness and versatility of the human voice originate from a complex interplay between the different parts in the vocal apparatus. The lungs and the oscillating vocal folds within the larynx mainly control pitch and loudness of a sound, whereas resonances and modulations in the vocal tract influence the timbre of a sound. During talking or singing, vibrations of the larynx can be recorded by throat microphones attached to the skin of the throat (s. Fig. 1). Such microphones typically use electret or piezo pick-ups to sense vibrations through contact with solid objects.

Capturing the human voice directly from the throat skin is advantageous since the recorded signals are not interfered by other sounds carried by the air. For this reason, throat microphones are used for communication in high-noise environments (e.g. by military and security agencies) and for speech health monitoring. Furthermore, because of their simple usage and robustness, the microphones are ideal for mobile and outdoor use. There are some disadvantages of throat microphones as well. Due to the missing contributions of the vocal tract, the recorded signals sound unnatural and muffled. In addition, the signal quality can be affected by tissue characteristics and facial hair on the user's throat. Some singers also complain about unpleasant pressure on the larynx during singing. Currently, we are experimenting with different sensor types and attachment mechanisms that may fix these issues. In the next section, we present an application scenario of throat microphones.

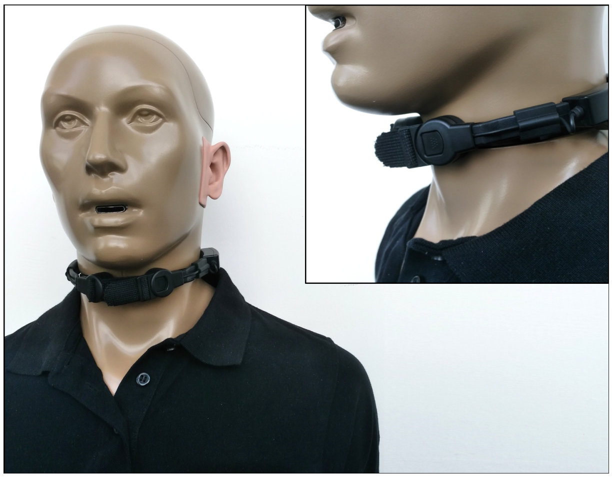

Figure 1. Throat microphone attached to a dummy head.

2. GEORGIAN MUSIC RESEARCH

Georgia is a country located in the Caucasus region of Eurasia with a rich polyphonic singing tradition. The centuries-old traditional three-voice chants have been acknowledged as intangible cultural heritage by the UNESCO in 2001. Recorded audio material plays an important role for research on Georgian vocal music, since transcriptions into Western staff notation do not appropriately capture the tuning and the tonal structure of this music. In this contribution, we rely on multitrack field record-

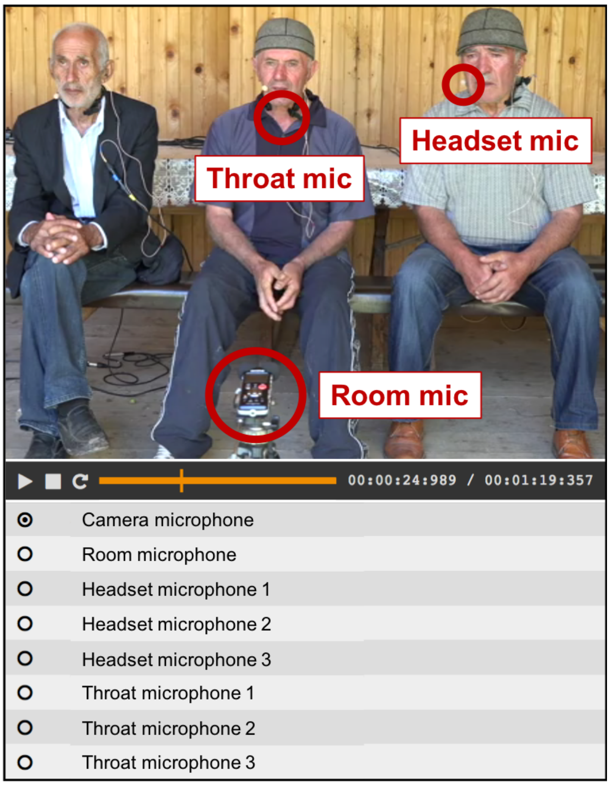

Figure 2. Multitrack audio/video player interface. Each singer is equipped with a throat microphone and a headset microphone. Additionally, the scene is recorded with a room microphone and a video camera.

ings of five Georgian songs [4]. In order to record the singers in their natural environment while being able to capture the singers' voices individually, each singer was equipped with a throat microphone and a headset microphone. Additionally, the singers were recorded with a video camera and a portable audio recorder ("room microphone"). The multitrack data is hosted on a publicly available website [1] using the *trackswitch.js*-functionality [5] which allows seamless switching and mixing of the individual tracks via an interactive user interface (s. Fig. 2). The player also invites listeners to discover the differences between throat and headset microphone recordings.

The analysis of the Georgian tuning system is part art of ongoing ethnomusicological research on Georgian music. In a study on historical tape recordings of three-voice chants, the authors annotated fundamental frequency (F0) trajectories of the three voices using a semi-automatic tool [1]. Based on the annotated trajectories, the authors detected high occurences of fifth and neutral third intervals in the three-voice chants. However, annotating large recording collections is time consuming and labor intensive. Throat microphone recordings can simplify such analyses since the required F0 trajectories can be estimated directly from the monophonic recordings using state-of-the-art monophonic pitch trackers. Experiments with F0 trajectories from recent multitrack throat microphone recordings confirm the findings in the historical recordings [3].

Besides their benefits for ethnomusicological research, throat microphone recordings constitute a useful basis for developing and evaluating mono- and multipitch estimation algorithms. F0 estimations in monophonic audio recordings are of limited accuracy, often requiring manual corrections or special analysis/synthesis approaches [2] in order to use the trajectories as reference annotations for evaluation purposes. First experiments show that due to the predominant pitch of the recorded voice, pitch trackers deliver a reliable performance on throat microphone recordings. This indicates that extracted F0 trajectories from throat microphone signals can serve as reference annotations without the need for further manual corrections.

[1] https://www.audiolabs-erlangen.de/resources/MIR/2018-ISMIR-LBD-ThroatMics

3. ACKNOWLEDGMENTS

The International Audio Laboratories Erlangen are a joint institution of the Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU) and Fraunhofer Institute for Integrated Circuits IIS. This work was supported by the German Research Foundation (DFG MU 2686/13-1, DFG SCHE 280/20-1). We would like to thank El Mehdi Lemnaouar for helping us with the website set-up.

4. REFERENCES

[1] Meinard Müller, Sebastian Rosenzweig, Jonathan Driedger, and Frank Scherbaum. Interactive fundamental frequency estimation with applications to ethnomusicological research. In *Proceedings of the AES International Conference on Semantic Audio*, pages 186–193, Erlangen, Germany, 2017.

[2] Justin Salamon, Rachel M. Bittner, Jordi Bonada, Juan José Bosch Vicente, Emilia Gómez Gutiérrez, and Juan P. Bello. An analysis/synthesis framework for automatic f0 annotation of multitrack datasets. In *Proceedings of International Society for Music Information Retrieval Conference (ISMIR)*, pages 71–78, Suzhou, China, 2017.

[3] Frank Scherbaum. On the benefit of larynx-microphone field recordings for the documentation and analysis of polyphonic vocal music. In *Proceedings of the International Workshop Folk Music Analysis*, pages 80–87, Dublin, Ireland, 2016.

[4] Frank Scherbaum, Nana Mzhavanadze, and Elguja Dadunashvili. A web-based, long-term archive of audio, video, and larynx-microphone field recordings of traditional Georgian singing, praying and lamenting with special emphasis on Svaneti. *International Symposium on Traditional Polyphony*, 2018, to appear.

[5] Nils Werner, Stefan Balke, Fabian-Robert Stöter, Meinard Müller, and Bernd Edler. trackswitch.js: A versatile web-based audio player for presenting scientific results. In *Proceedings of the Web Audio Conference (WAC)*, London, UK, 2017.