TILES Audio Recorder: An unobtrusive wearable solution to track audio activity

Tiantian Feng
Signal Analysis and Interpretation
Lab, University of Southern California
Los Angeles, California
tiantiaf@usc.edu

Amrutha Nadarajan
Signal Analysis and Interpretation
Lab, University of Southern California
Los Angeles, California
nadaraja@usc.edu

Colin Vaz
Signal Analysis and Interpretation
Lab, University of Southern California
Los Angeles, California
cvaz@usc.edu

Brandon Booth
Signal Analysis and Interpretation
Lab, University of Southern California
Los Angeles, California
bbooth@usc.edu

Shrikanth Narayanan
Signal Analysis and Interpretation
Lab, University of Southern California
Los Angeles, California
shri@sipi.usc.edu

ABSTRACT

Most existing speech activity trackers used in human subject studies are bulky, record raw audio content which invades participant privacy, have complicated hardware and non-customizable software, and are too expensive for large-scale deployment. The present effort seeks to overcome these challenges by proposing the TILES Audio Recorder (TAR) - an unobtrusive and scalable solution to track audio activity using an affordable miniature mobile device with an open-source app. For this recorder, we make use of Jelly Pro Mobile, a pocket-sized Android smartphone, and employ two open-source toolkits: openSMILE and Tarsos-DSP. Tarsos-DSP provides a Voice Activity Detection capability that triggers openSMILE to extract and save audio *features* only when the subject is speaking. Experiments show that performing feature extraction only during speech segments greatly increases battery life, enabling the subject to wear the recorder up to 10 hours at time. Furthermore, recording experiments with ground-truth clean speech show minimal distortion of the recorded features, as measured by root mean-square error and cosine distance. The TAR app further provides subjects with a simple user interface that allows them to both pause feature extraction at any time and also easily upload data to a remote server.

CCS CONCEPTS

• **Human-centered computing** → *Ubiquitous and mobile computing design and evaluation methods*;

KEYWORDS

Audio, wearable sensing, audio processing, openSMILE, privacy, human subjects study, audio feature recorder

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

WearSys '18, June 10, 2018, Munich, Germany
© 2018 Association for Computing Machinery.
ACM ISBN 978-1-4503-5842-2/18/06...\$15.00
<https://doi.org/10.1145/3211960.3211975>

ACM Reference Format:

Tiantian Feng, Amrutha Nadarajan, Colin Vaz, Brandon Booth, and Shrikanth Narayanan. 2018. TILES Audio Recorder: An unobtrusive wearable solution to track audio activity. In *WearSys '18: 4th ACM Workshop on Wearable Systems and Applications, June 10, 2018, Munich, Germany*. ACM, New York, NY, USA, Article 4, 6 pages. <https://doi.org/10.1145/3211960.3211975>

1 INTRODUCTION

Rapid advances in signal acquisition and sensing technology have produced miniaturized and low-power sensors capable of monitoring people's activities, physiological functions, and their environment [16]. When coupled with recent large-scale data analysis techniques, they show great promise for many applications, such as health monitoring, rehabilitation, fitness, and well being [14]. These advances have been primarily enabled by the ubiquitous mobile phone, which facilitates sensor setup, data transfer, and user feedback [11]. These non-invasive devices have made the design and implementation of medical, psychological, and behavioral research studies easier by reducing participant burden. We focus on applications of these technologies to audio recording of human speech in the wild.

Audio recording devices have come a long way, from bulky phonographs to hand-held microphones and lapel mics. The legacy systems are cumbersome, obtrusive, expensive, and not scalable. They can also potentially invade privacy by recording and storing raw audio. Mobile phones, laptops, and wearables are becoming a part of our everyday life. By using the miniature mics on these devices, a digital recorder [15][13][9] can overcome some of the drawbacks of the legacy systems, like obtrusiveness, and poor scalability. These modern devices are also capable of capturing day-long audio samples in naturalistic environments. Most digital recording solutions, like the EAR [15], afford privacy by allowing the user to delete recordings retrospectively. This process demands additional effort from the user to ensure their privacy is not violated and potentially jeopardizes the privacy of the people the user interacts with.

A recently-developed audio recording solution for human subjects studies is EAR [15]. EAR is an app that runs on a personal smart phone and records small fractions of audio through the day at regular intervals. The recording scheme senses audio only 5%

of the day and does not have an intelligent recording trigger to initiate the recording. This could potentially lead to recordings containing no audio activity or failing to record salient audio activity. Other notable efforts in the development of wearable audio recorders include devices like the Sociometer [7]. This device contains a microphone, motion sensor, and proximity sensor embedded in a shoulder pad that can be strapped on. However, this device is not commercially available which makes large-scale deployment prohibitive. Soundsense[13] is another audio wearable device that classifies different auditory scenes (music, speech, ambient sound) and events (gender, music, speech analysis) surrounding the user. It runs on mobile devices so it could scale to large study populations, but the platform is not publicly available. Unfortunately, recording solutions using phones usually suffer from poor recording quality because people commonly carry phones in their pockets or handbags. A study by VanDam [19] has shown that denser material composition and indirect microphone placement introduces amplitude loss at frequencies higher than 1 kHz.

In this paper, we introduce a new digital recording system called TILES¹ Audio Recorder (TAR), which enables passive sensing of audio activity in sensitive environments where recording raw audio is not a viable option. The recorder comes with an audio activity detector and a customizable recording scheme to initiate audio feature extraction as required by the demands of the investigator. The recorded features are not audible, but a large number of studies have successfully predict the emotions or stress from the speech features [5].

In the following sections, we give an overview of the TAR system and provide experimental results about its expected battery life and robustness of audio activity detection, and we also analyze the audio feature degradation at various distances from an audio source. Our results show that the system works as intended and introduces negligible feature degradation.

2 TAR OVERVIEW

In this section, we provide a detailed description of the hardware and software comprising the TAR system. TAR is a combination of a light-weight Android mobile device and an Android app to record audio features.

2.1 Motivation

As we described in Section 1, user privacy and unobtrusiveness are critical factors in designing a wearable audio recorder. In this study, we propose TAR, which is especially useful for deploying in environments like a hospital where privacy protections restrict raw audio recording. Moreover, TAR is designed to be unobtrusive and comfortable when worn. In addition, investigators can easily reproduce the setup given that TAR runs on consumer hardware and the software is freely available on Github. TAR introduces various features to address many study design challenges when using wearable audio recorders:²

- affordable cost allows scaling to a large number of participants

Figure 1: A recommended setup for TAR

- ease of placing the sensor close to the wearer's mouth provides consistently high recording quality
- extracting and storing audio features and immediately discarding raw audio protects user privacy
- an intuitive user interface enables wireless data transfer and data capture to be paused and resumed at any time

2.2 Hardware

TAR is designed to run on Android platforms because it is open source and runs on a variety of devices including small, lightweight, budget-friendly portable smartphones like the Jelly Pro [2]. The Jelly Pro smartphone, used in this work, has the following specifications:

- Dimensions and weight: 3.07x4.92x1.97 in, 60.4 g
- Processor: Quad-core processor, 1.1 GHz
- Battery capacity: 950 mAh
- Memory: 2 GB RAM / 16 GB ROM
- Sensors: microphone, camera, compass, gyroscope

A complete hardware setup consists of a Jelly Pro phone, a phone case, and a metal badge clip with PVC straps. The phone case is commercially available, and Jelly Pro is packed into the case. A metal badge clip, which is fastened to the case, is clipped to a user's clothes near the collar. With this setup, the distance from the wearer's mouth to the phone's integrated microphone typically ranges from 15 cm to 30 cm. A recommended setup is shown in Fig. 1.

2.3 Software

In this section, we describe the TAR Android application and the various service components and open source tools that enable recording of audio information.

App services The TAR Android application starts when the mobile phone switches on. The TAR app includes the following services running in the background:

- a Voice Activity Detection (VAD) service that runs periodically to listen for speech and if detected it triggers the audio feature extraction service
- an audio feature extraction service to extract audio features from the raw audio stream. The raw audio is immediately deleted after the features are extracted and only the features are saved.
- a battery service to record the battery percentage information every 5 minutes. The function of the service is to optimize the battery performance.

¹TILES: Tracking Individual pErformance using Sensors

²TAR github: <https://github.com/tiantiaf0627/TAR>

Figure 2: Recording control logic output with a simulated data stream

- a data upload service to securely transfer data to a remote server over WiFi.

Open-source libraries Most current mobile phone speech processing libraries rely on a frontend-backend architecture. The frontend running on the phone typically record and transmit audio snippets to the backend over an Internet connection. The backend, which commonly runs on a remote server, retrieves the audio recordings, and performs feature extraction. Such a solution reduces computational burden on the phone and decouples data acquisition and processing. However, in our experience, fewer participants are willing to wear devices that attempt to transmit the raw audio. To allow TAR to be used where security and privacy are a concern, the TAR app runs real-time speech processing engines on the phone to avoid transmitting raw audio.

In our TAR application, we incorporate two real-time speech processing libraries: openSMILE [8] and TarsosDSP [17]. TarsosDSP is a Java library for audio processing, which outputs real-time audio energy and pitch. It serves as a straight-forward implementation of a VAD in TAR. OpenSMILE is a tool for extracting a wide range of features from audio signals. The extracted audio features from openSMILE have been used for classifying emotional states and speaker properties [8]. In our application, the feature extraction runs entirely on the mobile device and an Internet connection is not required. One other benefit TAR gains as a result of using openSMILE is that it is highly configurable and allows investigators to extract various combinations of audio features to suit different purposes. Besides, we have also tested other potential DSP libraries including Funf [4]. However, Funf only extracts FFT and MFCCs from the audio stream, thus researchers are constrained in acquiring other valuable features from the audio.

3 TAR RECORDING SCHEME

In this section, we describe the VAD-triggered recording scheme used by TAR. We propose a recording scheme which only attempts to record audio features when audio activity is detected in order to reduce storage requirements and increase battery life. We use TarsosDSP [17] to implement the VAD. The TarsosDSP outputs the audio energy data every 10 ms, which is then used to determine the presence of audio activity by TAR.

Figure 3: TAR Battery life vs T_{off}

3.1 Recording Parameters

The TAR initiates feature recording when the VAD service is active and the recording trigger is on. T_{on} specifies the duration that the VAD service is run, and T_{off} controls the length of the idle state between two consecutive VAD services. A larger T_{off} means a longer idle state where no VAD service take place. In addition, we monitor the audio energy output from Tarsos-DSP to ensure that the feature extraction happens only when there is some audio activity. Two additional parameters control when the recording trigger is set to on: h and Δ . h is set as a threshold for speech energy and Δ sets the minimum time that speech energy needs to stay above h to trigger feature extraction. Once triggered, TAR runs the audio feature extraction service for l seconds. In TAR, we also add the flexibility to enable audio feature recordings periodically even when VAD service failed to detect the speech in order to sample some environment information.

The recording parameters in TAR are tunable and should be adjusted to provide the highest resolution of audio samples for the duration required by the researcher. Figure 3 presents the average battery operating time of TAR for different T_{off} (second) values. We tested and measured the battery operating time for 20 different Jelly

Figure 4: Schematic of experimental setup for examining VAD accuracy and feature degradation where $d_1 = 15, d_2 = 20, d_3 = 25, d_4 = 30$ cm

Pro devices for each value of T_{off} with $h = -\infty$ so the VAD service triggered audio feature recording every time.

4 RESULTS

As described in Section 3, TAR uses a VAD to extract audio features only during regions of audio activity. As the device is intended for use in the wild, we tested the VAD accuracy in noisy scenarios as a function of h . We also inspect if the device introduces any quality degradation in the features extracted. The following subsections describe the experimental setup and discuss the results of the VAD accuracy and feature degradation experiments.

4.1 Experimental Setup

To test VAD accuracy and feature distortion, we proposed a recording setup to allow TAR to record speech amplified through a speaker. We synthetically created two sets of audio, one for testing VAD accuracy and the other for testing feature degradation. For the VAD experiment, we randomly chose 250 gender-balanced utterances from the TIMIT database [10]. We concatenated all utterances and introduced silence of random duration (between 2 to 12 seconds) between utterances to simulate non-continuous speech. Additionally, we added noise from the DEMAND database [18] to the concatenated utterances at 10 dB, 5 dB, and 0 dB SNR levels to test the VAD performance in noisy speech. For the feature degradation experiment, we randomly sampled 1000 gender-balanced utterances from the TIMIT database and concatenated them into one file.

The audio files were then played through a loud-speaker. Multiple TARs set at distances 15 cm, 20 cm, 25 cm, and 30 cm from the speaker recorded the audio playback simultaneously. We chose these distances to mimic the typical range of distances the TAR would be from a user's mouth during use. Figure 4 shows a diagram of the experimental setup. For the VAD accuracy experiment, we chose to store outputs from the Tarsos-DSP and later set different thresholds for h and Δ on the stored output. For the feature quality experiment, we modified the operation of TAR to record continuously. The recordings from the devices were compared to the features extracted from the raw audio file.

4.2 VAD Accuracy Experiment

We set TAR at a distance of 20 cm from the speaker in the VAD accuracy experiment as it is a typical distance from a user's mouth. In order to evaluate our Tarsos-DSP VAD scheme, we first output a VAD result from the clean audio source file using the vadsohn function provided in the VOICEBOX speech processing toolbox [3]. The VAD result is a binary decision on a frame-by-frame basis,

Figure 5: VAD prediction accuracy as a function of energy threshold h using Tarsos-DSP library

TarsosDSP VAD Label	Baseline VAD Label		Source
	1	0	
	1	0	
1	83.63	16.37	Source
0	16.25	83.75	Source

TarsosDSP VAD Label	Baseline VAD Label		SNR=10dB
	1	0	
	1	0	
1	83.53	16.47	SNR=10dB
0	16.68	83.32	SNR=10dB

TarsosDSP VAD Label	Baseline VAD Label		SNR=5dB
	1	0	
	1	0	
1	82.20	17.80	SNR=5dB
0	21.72	78.28	SNR=5dB

TarsosDSP VAD Label	Baseline VAD Label		SNR=0dB
	1	0	
	1	0	
1	78.86	21.14	SNR=0dB
0	31.02	68.98	SNR=0dB

Figure 6: Confusion matrix of VAD output (%) for using Tarsos-DSP library and baseline output, where $h = -56$ dB

where each frame is 10 ms in length. Following this, we manually correct the vadsohn output in the regions that gives false predictions to produce a baseline VAD output. Meanwhile, we use the Tarsos-DSP output to generate the VAD decisions on the recorded audio in the clean and noisy conditions. As we stated in Section 2, VAD output is especially sensitive to h . To choose the best h , we vary h through -65 dB to -50 dB by 1 dB with Δ fixed at 100 ms. Finally, we compare the baseline VAD output and Tarsos-DSP VAD outputs at each h .

Figure 5 presents the VAD prediction accuracy in percentage with respect to h . We observe that VAD accuracy rates initially increase with h and then decrease significantly. Difference in the prediction accuracy among four speech SNR conditions is less significant when $h < -64$ dB or $h > -58$ dB. The prediction accuracy reaches the peak at -56 dB with clean speech, -56 dB with SNR=10 dB, -55 dB with SNR=5 dB and -54 dB with SNR=0 dB. Figure 6 presents the confusion matrix of VAD outputs between Tarsos-DSP and baseline with $h = -56$ dB. We observe that in clean speech and high SNR (10 dB, 5 dB) conditions, the negative predictive rate is above 75%, while in poor SNR (0 dB) condition, it drops to 68.98%. In addition, we note that the precision rates vary slightly ($< 5\%$) across SNR conditions. We advise users to setup similar recording experiments to calibrate h for own usage.

4.3 Feature Quality Experiment

In this experiment, we apply the emobase config setting [1] to extract the openSMILE low-level descriptors (LLDs) from the ground

Table 1: Deviation of LLDs in utterance-region

RMSE (root-mean-squared error)				
LLD	15cm	20cm	25cm	30cm
F_0 (Hz)	66.99	55.86	66.33	65.65
F_0 Env(Hz)	45.87	41.10	51.46	45.13
Loudness	0.863	0.507	0.352	0.352
MFCC[1-14]	45.52	43.77	41.45	42.81
LSP[1-8]	0.551	0.492	0.444	0.479
Cosine Distance				
LLD	15cm	20cm	25cm	30cm
MFCC[1-14]	0.199	0.172	0.175	0.184
LSP[1-8]	0.00195	0.00153	0.00139	0.00150

truth concatenated TIMIT file and recordings. The feature set contains 26 acoustic Low-Level Descriptors(LLDs), such as voicing related features(Pitch, voicing probability), energies (Intensity, Loudness), zero-crossing rate(ZCR) and cepstral information(MFCCs 1-14). The proposed LLDs are extracted every 10 ms using a 25 ms Hamming window

We compared the LLDs by measuring the root-mean-squared error (RMSE) and cosine distance between the features extracted from the source. Some features, however, are sensitive to energy levels, and they often show higher degradation in silence regions. Thus, we decided to compare the feature values only in utterance regions. Firstly, we perform the sanity check on the raw signals. We use the vadsohn function to remove signals in the silence regions, and we obtain the RMSE between source audio signal and recorded signal in utterance regions are 0.1297 (15 cm), 0.0790 (20 cm), 0.0590 (25 cm), and 0.0598 (30 cm).

We then use the vadsohn function to remove LLDs in the silence regions. RMSE and cosine distance between ground truth LLDs and TAR recorded LLDs in all test cases are listed in Table 1. Loudness, mel-frequency cepstral coefficients (MFCC) and line spectral pair (LSP) give largest RMSE in TAR#1 (15 cm) and lowest RMSE in TAR#3 (25 cm). Loudness is extremely sensitive to recording distance, where the maximum and minimum RMSE values are 0.863 and 0.352, respectively. RMSE in most LLDs initially decreases with recording distance, but then start to increase when the recording distance exceed 25 cm. Meanwhile, we observed that MFCC and LSP presents relatively small variations with recording distance in cosine distance. The deviations in LLDs comply with our sanity check results, where TAR#3 (25 cm) gains minimum degradation both in waveform signals and LLDs.

We plot the histograms of RMSE for F_0 , Loudness and histograms of cosine distance for MFCC in Fig 7. We perform the Kruskal-Wallis test [12] to investigate the consistence of RMSE among four recording distances. We find that the difference is significant in loudness ($p = 0.0018$), but it is fairly consistent in F_0 ($p = 0.3607$) and F_0 Env ($p = 0.2598$). MFCC exhibits most consistent measurement in both RMSE ($p = 0.8550$) and cosine distance ($p = 0.6005$). The results confirms that energy-related LLDs are sensitive to recording distance, but pitch and spectral LLDs yield consistent patterns with different recording distances. We also observe that majority errors of F_0 resides under 10 Hz, which confirms the robustness of the feature recorded by TAR.

Figure 7: (a), (b) presents the F_0 -RMSE distribution and the loudness-RMSE distribution, (c) displays MFCC-cosine distance distribution

5 USER EXPERIENCE

Fig.8 exhibits the GUI of TAR application. We programmed the GUI with two buttons for usage simplicity. The function of the top button is to allow users to quickly disable the audio recording service for privacy considerations. The other button presented in the GUI activates the uploading of collected audio features to the back-end server.

We tested TAR on 20 participants to assess user interface satisfaction. The participants were asked to wear and use the TAR for a minimum two hours. The Questionnaire for User Interface Satisfaction (QUIS), an instrument based on [6] was used to rate the TAR interface in terms of overall satisfaction to the software, screens, ease of learning to use the interface and system capabilities. TAR performs satisfactorily in all these aspects. Regarding the overall reaction to the software, an average rating of 7.8, 8.6 and 7.9 was reported on the 10 point Likert scale of Terrible/Wonderful, Difficult/Easy and Frustrating/Satisfying respectively. Ease of learning to use the system was reported at 9 and straightforwardness of performing tasks at 9. These numbers show that the intended test cases felt the UI was easy, simple and straightforward to use.

6 CONCLUSIONS

In this paper, we present TAR, an unobtrusive wearable solution to track audio activity in the wild. We described hardware and software comprising the TAR system. We then explained the VAD-trigger recording scheme in TAR Android App. We designed two experiments in examining the accuracy of proposed VAD scheme as well as the quality of the extracted audio features. The results

Figure 8: TAR Android App GUI. (a) UI display when TAR is idle and not recording. (b) UI display when TAR recording is turned on.

confirm the reliability of the VAD scheme. We observe minimal distortion of the recorded features except energy LLDs as measured by root-mean-square error and cosine distance. We also provided TAR to 20 participants, and we received high user satisfaction with the simple user interface. We believe the unobtrusiveness and flexibility of TAR, coupled with privacy-focused data acquisition, makes it suitable for a wide range of people-centric audio sensing applications.

Specifically, within next four months, we plan to deploy current TAR system over 150 volunteers who work at the USC Hospital. We intend to collect only the audio features from individual volunteer during working shift. We plan to use the collected audio features to investigate the intercommunication behavior, individual performance, and stress level of participants.

In future works, we plan to add the functions to automatically learn recording parameter instead of hand-tuning. We also aim to bring online analysis of the collected feature to classify contextual information or wearer's emotion variation in the future updates. At last, we plan to put the TAR Audio App on the Google Play store which is publicly accessible.

ACKNOWLEDGMENTS

The research is based upon work supported by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), via IARPA Contract No 2017-17042800005. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the ODNI, IARPA, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon.

REFERENCES

- [1] [n. d.]. emobase. <https://github.com/naxingyu/opensmile/blob/master/config/emobase.conf>
- [2] [n. d.]. JELLY. www.unihertz.com/jelly.html
- [3] [n. d.]. VOICEBOX: Speech Processing Toolbox for MATLAB. <http://www.ee.ic.ac.uk/hp/staff/dmb/voicebox/voicebox.html>
- [4] Nadav Aharony, Wei Pan, Cory Ip, Inas Khayal, and Alex Pentland. 2011. Social fMRI: Investigating and shaping social mechanisms in the real world. *Pervasive and Mobile Computing* 7 (2011), 643–659.
- [5] Carlos Busso, Zhigang Deng, Serdar Yildirim, Murtaza Bulut, Chul Min Lee, Abe Kazemzadeh, Sungbok Lee, Ulrich Neumann, and Shrikanth Narayanan. 2004. Analysis of Emotion Recognition Using Facial Expressions, Speech and Multimodal Information. In *Proceedings of the 6th International Conference on Multimodal Interfaces (ICMI '04)*. ACM, New York, NY, USA, 205–211.
- [6] John P Chin, Virginia A Diehl, and Kent L Norman. 1988. Development of an instrument measuring user satisfaction of the human-computer interface. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 213–218.
- [7] Tanzeem Choudhury and Alex Pentland. 2002. The Sociometer: A Wearable Device for Understanding Human Networks (*CSCW 02 Workshop*).
- [8] Florian Eyben, Martin Wöllmer, and Björn Schuller. 2010. Opensmile: The Munich Versatile and Fast Open-source Audio Feature Extractor. In *Proceedings of 18th ACM International Conference on Multimedia*. ACM, New York, NY, USA, 1459–1462.
- [9] Hillary Ganek and Alice Eriks-Brophy. 2018. Language ENvironment analysis (LENA) system investigation of day long recordings in children: A literature review. *Journal of Communication Disorders* 72 (2018), 77 – 85.
- [10] J. S. Garofolo, L. F. Lamel, W. M. Fisher, J. G. Fiscus, and D. S. Pallett. 1993. DARPA TIMIT acoustic-phonetic continuous speech corpus CD-ROM. NIST speech disc 1-1.1. *NASA STI/Recon Technical Report N 93* (Feb. 1993).
- [11] T. Klingenberg and M. Schilling. 2012. Mobile wearable device for long term monitoring of vital signs. *Computer Methods and Programs in Biomedicine* 106, 2 (2012), 89 – 96.
- [12] William H. Kruskal and W. Allen Wallis. 1952. Use of Ranks in One-Criterion Variance Analysis. *J. Amer. Statist. Assoc.* 47, 260 (1952), 583–621.
- [13] Yuhao Liu, James J. S. Norton, Raza Qazi, Zhanan Zou, Kaitlyn R. Ammann, Hank Liu, Lingqing Yan, Phat L. Tran, Kyung-In Jang, Jung Woo Lee, Douglas Zhang, Kristopher A. Kilian, Sung Hee Jung, Timothy Bretl, Jianliang Xiao, Marvin J. Slepian, Yonggang Huang, Jae-Woong Jeong, and John A. Rogers. 2016. Epidermal mechano-acoustic sensing electronics for cardiovascular diagnostics and human-machine interfaces. *Science Advances* 2, 11 (2016).
- [14] T. Martin, E. Jovanov, and D. Raskovic. 2000. Issues in wearable computing for medical monitoring applications: a case study of a wearable ECG monitoring device. In *Digest of Papers. Fourth International Symposium on Wearable Computers*. 43–49.
- [15] Matthias R. Mehl, James W. Pennebaker, D. Michael Crow, James Dabbs, and John H. Price. 2001. The Electronically Activated Recorder (EAR): A device for sampling naturalistic daily activities and conversations. *Behavior Research Methods, Instruments, Computers* 33, 4 (01 Nov 2001), 517–523.
- [16] Shyamal Patel, Hyung Park, Paolo Bonato, Leighton Chan, and Mary Rodgers. 2012. A review of wearable sensors and systems with application in rehabilitation. *Journal of NeuroEngineering and Rehabilitation* 9, 1 (20 Apr 2012), 21.
- [17] Joren Six, Olmo Cornelis, and Marc Leman. 2014. TarsosDSP, a Real-Time Audio Processing Framework in Java. In *Proceedings of the 53rd AES Conference (AES 53rd)*.
- [18] Joachim Thiemeann, Nobutaka Ito, and Emmanuel Vincent. 2013. The diverse environments multi-channel acoustic noise database: A database of multichannel environmental noise recordings. *Journal of the Acoustical Society of America* 133, 5 (2013), 3591–3591.
- [19] Mark VanDam. 2014. Acoustic characteristics of the clothes used for a wearable recording device. *The Journal of the Acoustical Society of America* 136, 4 (2014), EL263–EL267.