SkullConduct: Biometric User Identification on Eyewear Computers Using Bone Conduction Through the Skull

Stefan Schneegass
Human-Computer Interaction Group
University of Stuttgart
stefan.schneegass@vis.uni-stuttgart.de

Youssef Oualil
Spoken Language Systems Group
Saarland University
youssef.oualil@lsv.uni-saarland.de

Andreas Bulling
Perceptual User Interfaces Group
Max Planck Institute for Informatics
bulling@mpi-inf.mpg.de

ABSTRACT
Secure user identification is important for the increasing number of eyewear computers but limited input capabilities pose significant usability challenges for established knowledge-based schemes, such as passwords or PINs. We present *SkullConduct*, a biometric system that uses bone conduction of sound through the user's skull as well as a microphone readily integrated into many of these devices, such as Google Glass. At the core of SkullConduct is a method to analyze the characteristic frequency response created by the user's skull using a combination of Mel Frequency Cepstral Coefficient (MFCC) features as well as a computationally light-weight 1NN classifier. We report on a controlled experiment with 10 participants that shows that this frequency response is person-specific and stable – even when taking off and putting on the device multiple times – and thus serves as a robust biometric. We show that our method can identify users with 97.0% accuracy and authenticate them with an equal error rate of 6.9%, thereby bringing biometric user identification to eyewear computers equipped with bone conduction technology.

Author Keywords
User Authentication; User Identification; Bone Conduction; Eyewear Computer; Google Glass

ACM Classification Keywords
H.5.m. Information Interfaces and Presentation(e.g. HCI): Miscellaneous; K.6.5 Computing Milieux: Security and Protection: Authentication

INTRODUCTION
Secure user authentication is important for personal devices, such as mobile phones, given that these devices store an increasing amount of personal information. To address limitations of established knowledge-based authentication schemes, such as passwords and PINs, recent works exploit the sensors readily integrated into these devices. For example, previous works proposed the analysis of keystroke dynamics [12], gait patterns [13], ambient sound [10], micro-movements while interacting [2], the shape of the user's ear [8], bioimpedance [5], or the way a user places or answers a phone call [4].

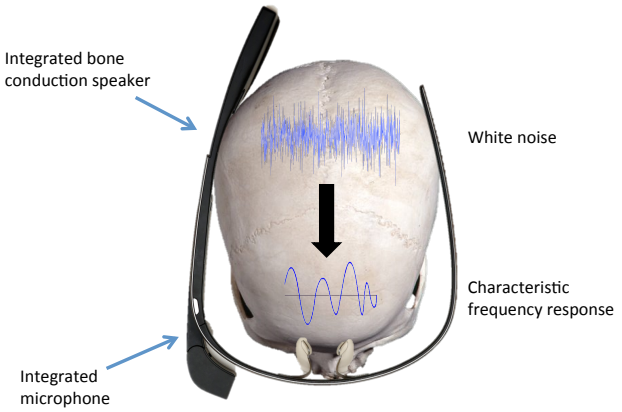

Figure 1. SkullConduct uses the bone conduction speaker and microphone readily integrated into the eyewear computer and analyses the characteristic frequency response of an audio signal sent through the user's skull.

In contrast, secure user authentication on another type of personal device, namely eyewear computers, remains largely unexplored. Simkin et al. described an approach to use Google Glass in combination with challenge-response protocols for authentication with an external system, such as an ATM or an entrance door [21]. The lack of methods to authenticate with eyewear computers themselves is partly because these devices only recently became widely available but also because their limited input capabilities and unique affordances pose usability challenges for traditional authentication schemes. Google glass for example uses the combination of strokes and taps on the touch-sensitive side of the device for authentication. Another notable exception is the recent work by Rogers et al. who explored the analysis of users' blinking patterns and head movements for user identification on Google Glass [19].

We present *SkullConduct*, a biometric system that uses bone conduction of sound through the user's skull for secure user identification and authentication on eyewear computers. Bone conduction has been used before as a transmission concept in different consumer devices, such as hands-free head-

CHI'16, May 07 - 12, 2016, San Jose, CA, USA

ACM 978-1-4503-3362-7/16/05...$15.00
DOI: http://dx.doi.org/10.1145/2858036.2858152

sets and headphones, bone anchored hearing aids, as well as special-purpose communication systems, such as for diving or high-noise environments. Bone conduction has only recently become available on eyewear computers, such as Google Glass, as a privacy-preserving means of relaying information to the user. SkullConduct uses the microphone readily available on many of these devices to analyse the frequency response of the sound after it travelled through the user's skull (see Figure 1). As shown in this work, individual differences in skull anatomy result in highly person-specific frequency responses that can be used as a biometric.

The contributions of this work are two-fold. First, we present SkullConduct, a biometric system that uses bone conduction for secure user identification and authentication on eyewear computers. The system combines Mel Frequency Cepstral Coefficient (MFCC) based features with a light-weight 1NN classifier that can directly run on Google Glass. In contrast to established uses of bone conduction, our system uses a microphone close to the user's head – in our case readily integrated into Glass – to measure and analyze the frequency response after the sound traveled through the user's skull.

Second, we report on a controlled experiment with 10 participants that shows that this frequency response serves as a robust biometric, even when taking off and putting on the device multiple times. We show that we can identify users with 97.0% accuracy and authenticate them with an equal error rate of 6.9%.

THE SKULLCONDUCT SYSTEM

There are, in general, two different pathways audio can take to get from a source to the user. The most widely used pathway, as for example in the case of headphones or speakers, is via air conduction in which the audio travels through the air and the auditory channel to the user's inner ear. The second pathway is via bone conduction, i.e. directly through the skull to the inner ear. Especially for eyewear computers that already are located close or even at the head of the user, using bone conduction yields the advantage that the audio is not well audible to bystanders and thus more private.

So far, systems typically used speech to identify different users (see [18] for an example). In contrast, SkullConduct exploits the characteristic changes in an audio signal while it travels through a user's skull (see Figure 1). When audio is played back with a bone conduction speaker (i.e., the audio travels through the head) it is modified by the user's head. If recorded with a microphone, the changes in the audio signal reflect the specific characteristics of the user's head. Since the structure of the human head includes different parts such as the skull, tissues, cartilage, and fluids and the composition of these parts and their location differ between users, the modification of the sound wave differs between users as well. First, the speed of sound transmission differs for each of the parts of the human head [15] and, second, the different signal frequencies are damped differently [22]. In this work we opted for Gaussian white noise as the input signal since it covers the whole frequency range and therefore all frequency bands that might get affected by individual skull characteristics.

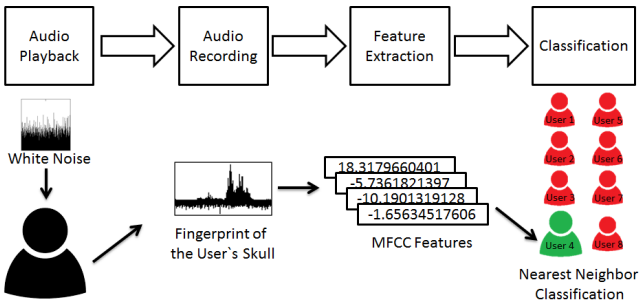

Figure 2. The recognition pipeline we used to authenticate users: (1) white noise is played back using the bone conduction speaker, (2) the user's skull influences the signal in a characteristic way, (3) MFCC features are extracted, and (4) a 1NN algorithm is used for classification.

We implemented SkullConduct on Google Glass as one of the most common used smart eyewear devices. We developed an application that plays back an audio file using the integrated bone conduction speaker and records concurrently with the integrated microphone. The recording is saved on the Glass as a byte file with 44100 samples per second, a single channel (i.e., mono), and a precision of two bytes per sample. To authenticate users, the system is capable of extracting features from the recording and comparing it to a training set using a 1NN classifier.

Recognition Pipeline

Our recognition pipeline to identify and authenticate users combines Mel Frequency Cepstral Coefficients (MFCCs) [6] as acoustic features with a computationally light-weight 1-nearest-neighbour classifier (see Figure 2). MFCCs are commonly used in speech classification and speaker identification but were shown to also perform well for non-speech event classification (see [17] for an example). In a first step, the signal as a whole is transformed using a Fourier transform. Afterwards, the power spectrum is mapped to the Mel scale using a Mel Filter Bank (MFB). Then, the Discrete Cosine Transform (DCT) is calculated after taking the logarithm. Finally, the MFCCs are given by the 2-13 DCT coefficients. In this work, we extend the 12 MFCCs features with their first derivatives (deltas) resulting in 24 features. All features are then used as input to a 1-nearest-neighbor classifier to identify or authenticate users. As a distance measure we used the sum of the Euclidean distances of each feature of a sample.

Application Scenarios

We envision two main application scenarios in which our system will be useful.

Personalization of Eyewear Computers

Eyewear computers are used in an increasing number of applications, such as for training in laboratories [9], medical documentation [1], educational purposes [11], or even during surgeries [14]. In all of these domains, multiple users may use a single device on a regular basis. As soon as a user puts on the device, SkullConduct can immediately identify the user and configure user-specific settings, such as preferred applications or system preferences.

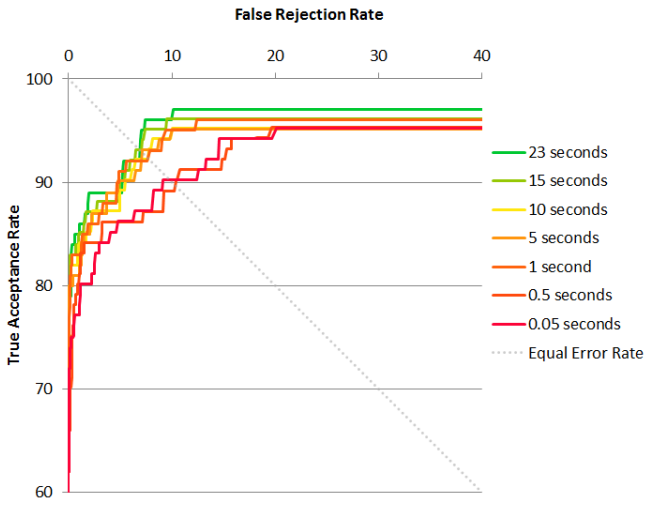

Figure 3. Receiver operating characteristic (ROC) curves summarizing the performance in terms of true acceptance rate vs. false acceptance rate for different recording lengths.

Protecting Private Content

Most eyewear computers are personal devices that contain sensible information about the owner, such as social media logins or bank account details. Current systems, such as Google Glass, are not protected and attackers can simply put on the device and access sensitive information. User authentication could automatically be triggered after the device has been put on by a new user. In addition, as soon as specific applications are started, such as the banking application, SkullConduct could re-authenticate the user to ensure he is allowed to access the application data.

EVALUATION

We evaluated SkullConduct with respect to the two main operating modes of biometric systems, namely user identification and authentication [16]. We designed a user study to record characteristic frequency responses for multiple people wearing Google Glass in a controlled laboratory setting.

Data Collection

We recorded data of 10 participants (9 male, 1 female) aged between 21 and 36 years ($M = 28$, $SD = 4.35$). The recording took place in a quiet room without any other source of noise and the participants sat down on a chair in the middle of the room. In this initial evaluation of the approach, we opted to have no confounding audio sources that may influence our results, such as sounds of other electronic devices or people. As mentioned before, we used a randomly generated Gaussian white noise audio signal with a length of 23 seconds. We recorded each participant 10 times with the same audio signal. After five recording trials, we asked participants to take off the device and put it back on to include different placements of the device on the participant's head.

Analysis

After recording the samples of all users, we analyzed the recorded data using 10-fold cross validation. In each fold, similar to Holz et al. [8], we excluded all recordings of one participant (i.e., the attacker). Within each fold, we did an additional two-fold cross validation. To this end, we grouped the recordings of the nine remaining participants into two folds. The first five recordings went into the first fold and the second five recordings, recorded after taking off Google Glass and putting it back on, went into the second fold. In total, we trained our system with 45 recordings (i.e., fold 1) of the nine known participants and used 55 recordings (i.e., fold 2, 45 recordings from known and 10 from unknown participants) for testing. We deliberately chose to split the data of each user since the placement of the Google Glass might influence the results [7, 22].

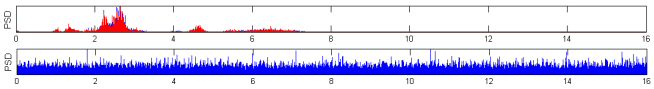

Figure 4. The user-specific modification (top) of the white noise input signal (bottom) takes place in the range of 1 kHz to 8 kHz with most modifications occurring between 2 kHz and 4 kHz.

User Identification

The first evaluation task for our SkullConduct system is to identify a known user. This might be necessary when an eyewear computer is shared within a group of users (i.e., a family or at work) but not require authentication since, for example, this information is only used for personalization of the device and not to protect content. In our case, the system achieves a 97.0% accuracy (cf., Figure 3 – True Acceptance Rate (TAR)). Thus, the lowest Euclidean distance between the new sample is with a recorded sample of the same user. In only 3% of the cases, a user is mistaken with another one.

User Authentication

The second evaluation task for our system is to authenticate a known user while rejecting unknown ones. The main measure of goodness for authentication system is the Equal Error Rate (EER) which is the point for which the false acceptance rate (FAR) and the false rejection rate (FRR) are equal [16]. We calculate both rates for our system for specific thresholds that decide whether a user will be authenticated or rejected (i.e., the euclidean distance between a training data and the authentication data needs to be lower than the threshold). Next, we calculated the EER out of the FAR and FRR (cf., [16]). The FAR is the percentage of samples that are mistakenly granted access even though they are from an unknown user. in contrast, the FRR is the percentage of samples that are mistakenly refused to access even though they are from known users. In our case, both rates were the same at 6.9%. The receiver operating characteristic (ROC) curve in Figure 3 shows the SkullConduct precision for different thresholds. For a high true authentication precision (97.0%), the FAR was 10.2%.

Influence of Different Frequency Bands

As related work suggested [22], different frequencies are influenced in characteristic ways by the head and skull. To investigate this phenomena, we calculated the Power Spectral Density (PSD), which describes how the power of the signal traversing the skull is distributed over the frequency range (see Figure 4). As can be seen from the figure, the head and skull for each participant influenced the PSD of the original signal in a specific way (see Figure 5). This influence varies

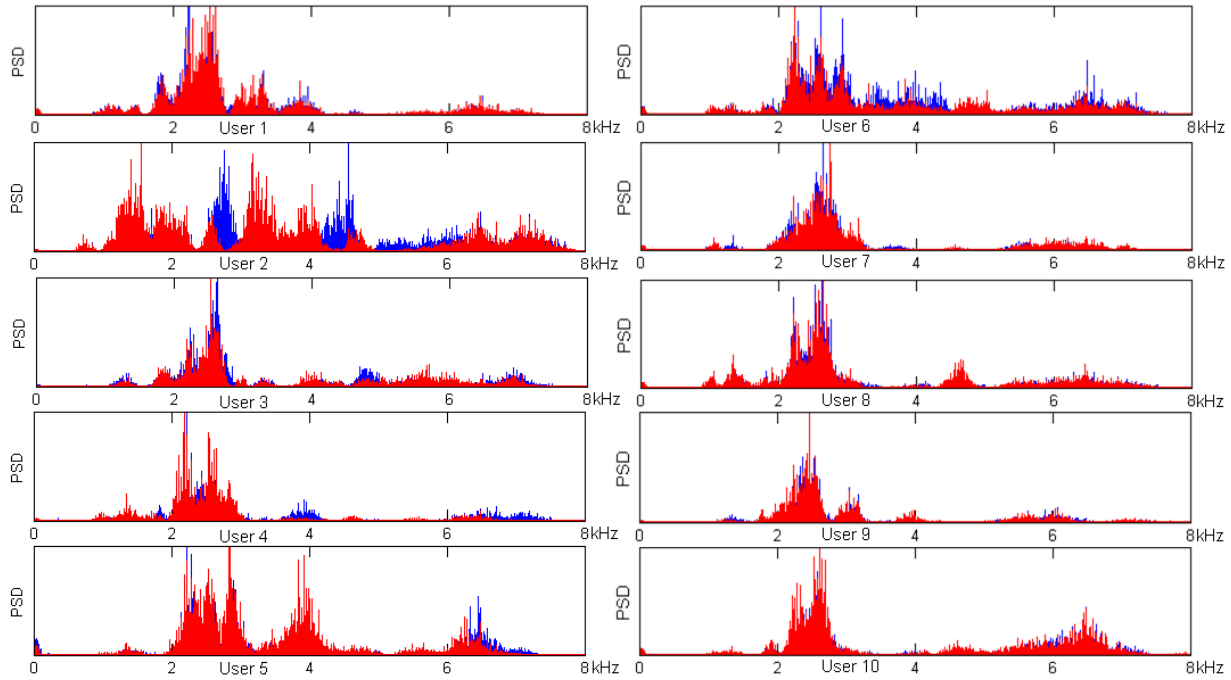

Figure 5. The power spectral density visualized for two recordings of each of the ten participants before and after removing and putting the device back on (trial 1 shown in red and trial 2 in blue) in the range of 0 kHz to 8 kHz. The changes in the power spectral density are almost similar for the different placements but differ between participants.

among participants but is constant for the same user even over several trails (i.e., only slightly affected by misplacement of the eyewear computer). Furthermore, the user-specific differences are skewed towards the lower frequency ranges and the main influence of the user's skull is for frequencies between 2 kHz and 4 kHz.

Influence of Audio Length

Current authentication systems on mobile devices require about 1.5 seconds to authenticate a user [20]. We used audio recordings of 23 seconds length which would take significantly longer for a user to authenticate. Therefore, we evaluated the performance of our system using audio with shorter lengths. Specifically, we cut each recording after 15, 10, 5, 1, 0.5, or 0.05 seconds and calculated a ROC curve for each length of audio samples using the same procedure as described before. As shown in Figure 3, the EER significantly drops when using audio samples shorter than 1 second.

DISCUSSION AND LIMITATIONS

The evaluation of our system yielded promising results. We showed that bone-conduction audio is well suited as a biometric security system for eyewear computers. However, we tested our approach only in a controlled setting without any background noise. Thus, we used a best-case scenario for our user study to explore the general feasibility of our approach. It will be interesting to see if and how much additional noise, such as other people talking in the room or appliances, reduces performance. One potential solution to this problem are algorithms that preserve the specific characteristics of each skull but remove the environmental influences [3]. Furthermore, there might be additional influences such as hair growth or gained weight that might impact the accuracy of our approach and need to be evaluated in the future. Although we show that a white noise signal of 1 second is sufficient to achieve high authentication accuracy, white noise signals may be unpleasant for the user. In the future, we envision that white noise could be replaced by more pleasant audio sounds such as common start-up jingles or even short music clips. Any alternative sound, however, needs to cover a sufficient number of frequency bands to discriminate well between different users.

CONCLUSION

We presented SkullConduct, a biometric system that exploits the characteristic frequency response of the human skull for user identification and authentication on eyewear computers equipped with bone conduction technology, such as Google Glass. While other biometric systems require the user to enter information explicitly (e.g., place the finger on a fingerprint reader), our system does not require any explicit user input. We implemented our system on Google Glass and evaluated its performance in a controlled user study with 10 participants. We demonstrated that our approach works well and can identify users with 97.0% accuracy as well as an EER of 6.9%.

ACKNOWLEDGMENTS

This work has in part been funded by the Cluster of Excellence on Multimodal Computing and Interaction (MMCI) at Saarland University, DLR Technology Marketing, the Helmholtz Validation Fund, the Deutsche Forschungsgemeinschaft (DFG) under grant SFB 1102, the European Union through the FP7 project Metalogue under grant agreement 611073, and the EU 7th Framework Programme under grant agreement no. 323849.

REFERENCES

1. Urs-Vito Albrecht, Ute von Jan, Joachim Kuebler, Christoph Zoeller, Martin Lacher, Oliver J Muensterer, Max Ettinger, Michael Klintschar, and Lars Hagemeier. 2014. Google Glass for documentation of medical findings: evaluation in forensic medicine. *Journal of medical Internet research* 16, 2 (2014).
2. Cheng Bo, Lan Zhang, Xiang-Yang Li, Qiuyuan Huang, and Yu Wang. 2013. SilentSense: Silent User Identification Via Touch and Movement Behavioral Biometrics. *Proceedings of the 19th annual international conference on Mobile computing & networking - MobiCom '13* (2013), 187. DOI: http://dx.doi.org/10.1145/2500423.2504572
3. Pedro Cano, Eloi Batlle, E Batle, T Kalker, and J Haitsma. 2002. A review of algorithms for audio fingerprinting. In *Multimedia Signal Processing, 2002 IEEE Workshop on*. 169–173. DOI: http://dx.doi.org/10.1109/MMSP.2002.1203274
4. M. Conti, I. Zachia-Zlatea, and Bruno Crispo. 2011. Mind how you answer me!: transparently authenticating the user of a smartphone when answering or placing a call. *Proceedings of the 6th ACM Symposium on Information, Computer and Communications Security* (2011), 249–259. DOI: http://dx.doi.org/10.1145/1966913.1966945
5. Cory Cornelius, Ronald Peterson, Joseph Skinner, Ryan Halter, and David Kotz. 2014. A Wearable System That Knows Who Wears It. In *Proceedings of the 12th Annual International Conference on Mobile Systems, Applications, and Services (MobiSys '14)*. ACM, New York, NY, USA, 55–67. DOI: http://dx.doi.org/10.1145/2594368.2594369
6. Steven B. Davis and Paul Mermelstein. 1980. Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. *IEEE Transactions on Acoustics, Speech, and Signal Processing* 28, 4 (1980). DOI: http://dx.doi.org/10.1109/TASSP.1980.1163420
7. Paula Henry and Tomasz R Letowski. 2007. Bone Conduction : Anatomy , Physiology , and Communication. May (2007).
8. Christian Holz, Senaka Buthpitiya, and Marius Knaust. 2015. Bodyprint: Biometric User Identification on Mobile Devices Using the Capacitive Touchscreen to Scan Body Parts. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 3011–3014. DOI: http://dx.doi.org/10.1145/2702123.2702518
9. Grace Hu, Lily Chen, Johanna Okerlund, and Orit Shaer. 2015. Exploring the Use of Google Glass in Wet Laboratories. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*. ACM, 2103–2108.
10. Nikolaos Karapanos, Claudio Marforio, Claudio Soriente, and Srdjan Čapkun. 2015. Sound-Proof: Usable Two-Factor Authentication Based on Ambient Sound. *arXiv:1503.03790 [cs]* (2015). http://arxiv.org/abs/1503.03790
11. Paul Lukowicz, Andreas Poxrucker, Jens Weppner, B. Bischke, Jochen Kuhn, and Michael Hirth. 2015. Glass-physics: Using Google Glass to Support High School Physics Experiments. In *Proceedings of the 2015 ACM International Symposium on Wearable Computers*. 151–154. DOI: http://dx.doi.org/10.1145/2802083.2808407
12. Emanuele Maiorana, Patrizio Campisi, Noelia González-Carballo, and Alessandro Neri. 2011. Keystroke dynamics authentication for mobile phones. *Proceedings of the 2011 ACM Symposium on Applied Computing SAC 11* (2011), 21–26. DOI: http://dx.doi.org/10.1145/1982185.1982190
13. Jani Mantyjarvi, Mikko Lindholm, Elena Vildjiounaite, Satu-Marja Makela, and Heikki Ailisto. 2005. Identifying users of portable devices from gait pattern with accelerometers. In *Acoustics, Speech, and Signal Processing, 2005. Proceedings. (ICASSP '05). IEEE International Conference on*, Vol. 2. ii/973–ii/976 Vol. 2. DOI: http://dx.doi.org/10.1109/ICASSP.2005.1415569
14. Oliver J Muensterer, Martin Lacher, Christoph Zoeller, Matthew Bronstein, and Joachim Kübler. 2014. Google Glass in pediatric surgery: An exploratory study. *International Journal of Surgery* 12, 4 (2014), 281–289.
15. William D. Jr. O'Brien and Y Liu. 2005. Evaluation of acoustic propagation paths into the human head. 2005 (2005), 1–24. http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=ADA437351
16. P. Jonathon Phillips, Alvin Martin, C. L. Wilson, and Mark Przybocki. 2000. Introduction to evaluating biometric systems. *Computer* 33, 2 (2000), 56–63. DOI: http://dx.doi.org/10.1109/2.820040
17. Axel Plinge, Rene Grzeszick, and Gernot A. Fink. 2014. A Bag-of-Features approach to acoustic event detection. In *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 3704–3708. DOI: http://dx.doi.org/10.1109/ICASSP.2014.6854293
18. Sandra Pruzansky. 1963. Pattern Matching Procedure for Automatic Talker Recognition. *The Journal of the Acoustical Society of America* 35, 3 (1963), 214–215.
19. Cynthia E Rogers, Alexander W Witt, Alexander D Solomon, and Krishna K Venkatasubramanian. 2015. An Approach for User Identification for Head-mounted Displays. In *Proceedings of the 2015 ACM International Symposium on Wearable Computers (ISWC '15)*. ACM, New York, NY, USA, 143–146. DOI: http://dx.doi.org/10.1145/2802083.2808391

20. Stefan Schneegass, Frank Steimle, Andreas Bulling, Florian Alt, and Albrecht Schmidt. 2014. SmudgeSafe: Geometric Image Transformations for Smudge-resistant User Authentication. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '14)*. ACM, New York, NY, USA, 775–786. DOI: http://dx.doi.org/10.1145/2632048.2636090

21. Mark Simkin, Dominique Schrder, Andreas Bulling, and Mario Fritz. 2014. Ubic: Bridging the Gap between Digital Cryptography and the Physical World. In *Proc. of the 19th European Symposium on Research in Computer Security (ESORICS 2014)* (2014-09-01), Vol. 8712. Springer International Publishing, 56–75. http://dx.doi.org/10.1007/978-3-319-11203-9_4https://perceptual.mpi-inf.mpg.de/files/2014/09/Simkin14_ubic1.pdf

22. Stefan Stenfelt and Richard L Goode. 2005. Transmission properties of bone conducted sound: measurements in cadaver heads. *The Journal of the Acoustical Society of America* 118, 4 (2005), 2373–2391. DOI:http://dx.doi.org/10.1121/1.2005847