HPSpeech: Silent Speech Interface for Commodity Headphones

Ruidong Zhang
rz379@cornell.edu
Cornell University
Ithaca, NY, USA

Hao Chen
hc732@cornell.edu
Cornell University
Ithaca, NY, USA

Devansh Agarwal
da398@cornell.edu
Cornell University
Ithaca, NY, USA

Richard Jin
rj284@cornell.edu
Cornell University
Ithaca, NY, USA

Ke Li
kl975@cornell.edu
Cornell University
Ithaca, NY, USA

François Guimbretière
fvg3@cornell.edu
Cornell University
Ithaca, NY, USA

Cheng Zhang
chengzhang@cornell.edu
Cornell University
Ithaca, NY, USA

ABSTRACT

We present HPSpeech, a silent speech interface for commodity headphones. HPSpeech utilizes the existing speakers of the headphones to emit inaudible acoustic signals. The movements of the temporomandibular joint (TMJ) during speech modify the reflection pattern of these signals, which are captured by a microphone positioned inside the headphones. To evaluate the performance of HPSpeech, we tested it on two headphones with a total of 18 participants. The results demonstrated that HPSpeech successfully recognized 8 popular silent speech commands for controlling the music player with an accuracy over 90%. While our tests use modified commodity hardware (both with and without active noise cancellation), our results show that sensing the movement of the TMJ could be as simple as a firmware update for ANC headsets which already include a microphone inside the hear cup. This leaves us to believe that this technique has great potential for rapid deployment in the near future. We further discuss the challenges that need to be addressed before deploying HPSpeech at scale.

CCS CONCEPTS

• **Human-centered computing → Ubiquitous and mobile computing systems and tools**; *Gestural input*; • **Computing methodologies** → Speech recognition.

KEYWORDS

Silent Speech; Acoustic Sensing; Headphones; Commodity-off-the-shelf (COTS)

ISWC '23, October 08–12, 2023, Cancun, Quintana Roo, Mexico
© 2023 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 979-8-4007-0199-3/23/10…$15.00
https://doi.org/10.1145/3594738.3611365

ACM Reference Format:
Ruidong Zhang, Hao Chen, Devansh Agarwal, Richard Jin, Ke Li, François Guimbretière, and Cheng Zhang. 2023. HPSpeech: Silent Speech Interface for Commodity Headphones. In *Proceedings of the 2023 International Symposium on Wearable Computers (ISWC '23), October 08–12, 2023, Cancun, Quintana Roo, Mexico.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3594738.3611365

1 INTRODUCTION

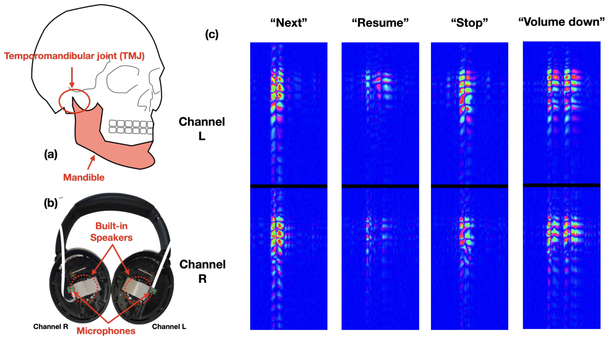

Figure 1: The HPSpeech system. (a) Illustration of sensing principles. HPSpeech uses active acoustic sensing to detect the movements of the TMJ to decode silent speech. (b) Illustration of the HPSpeech device: two miniature microphones are installed inside a pair of Bose QC 45 headphones. The built-in speakers are used. (c) Echo profiles for different utterances.

Headphones are among the most popular wearable devices on the consumer market. The primary purpose of headphones is to listen to music or watch videos using a music or video player. There are two main methods of interacting with headphones to input commands on these music/video players. The first method requires users to use their hands to operate the buttons on the headphones or their phone or laptop. It requires users to divert their attention from their current task and can be especially challenging when their hands are busy. The second interaction approach involves using speech commands to control the headphones. While speech input provides a hands-free interaction, it may not always be convenient in certain settings such as in a quiet library. In comparison to voice-based speech interaction, a silent speech interface (SSI) recognizes speech without the need for sound, eliminating the requirement of speaking out loud. Silent speech allows users to express their interaction intentions discreetly and in a hands-free manner, without disturbing the surrounding environment.

Recognizing silent speech has been a challenging task for the wearable community. Prior work has explored tracking the movements of tongue using magnetic sensors [1, 2, 7, 8, 16, 17, 26] or capactive sensors [18, 19, 23] inside the oral cavity, which is inconvenient for many users. Recent work also explored placing sensors on the skin around head to capture articulator movements during speech. For instance, researchers explored using ultrasonic imaging probes under the chin to directly "see" the tongue [5, 6, 13, 21, 29, 31] or using electromyography (EMG) to capture muscle movement-related electric signals [14, 15, 24, 25, 27, 30]. However, these sensing systems still require skin-contacting sensors or electrodes, which may not be comfortable or socially-acceptable. To further improve the level of comfort and social acceptability, researchers recently proposed many methods that do not require sensors at obvious locations, such as behind the ear [28], inside the ear canal [12], using existing form factors such as earphones/headphones [3], necklaces [20, 32], VR-headset [34] or glassframes [33]. However, they still require customized hardware or significant modification to existing form factors, which may not be immediately deployable on the commodity devices. The most recent work, EarCommand [12] use active acoustic sensing to capture the deformation inside ear canal to recognize silent speech. However, this approach does not apply to over-the-ears headphones which do not place speaker and microphones into ear canal. Despite the fact that many headphones are already equipped with a rich set of acoustic sensors (speakers and microphones), silent speech recognition on headphones has not been explored, partly due to the difficulty to capture enough useful information. To the best of our knowledge, HPSpeech is the first work to implement SSI by utilizing the acoustic sensors that are already embedded into many off-the-shelf headphones (especially with active noise cancellation).

In this paper, we present the design and implementation of HPSpeech, a silent speech interface for commodity headphones that can recognize 8 silent speech phrases to control music player. HPSpeech utilizes the existing speakers of the headphones to emit inaudible acoustic signals. The movements of the temporomandibular joint (TMJ) during speech alter the reflections of the signal before it is captured by a microphone positioned inside the headphone.

To understand the performance of HPSpeech, we evaluated it with 18 participants in total on two commodity headphones: Bose QC45 and Adesso Xtream G1. With a customized acoustic data processing and deep learning pipeline, HPSpeech was able to distinguish 8 popular silent speech commands to control a music player with over 90% accuracy, even when the speakers were playing music through the headphones during the entire study. Because HPSpeech utilizes the built-in speakers of commodity headphones and only needs a miniature microphone inside, it has a great potential to be deployed on millions of headphones in the near future. For instance, for devices with build-in microphones such as noise-cancelling headphones, it might be even possible to deploy HPSpeech with a firmware update. We further discuss the challenges needed to be addressed before it can be deployed at scale. The contributions of the paper are:

- We are the first to demonstrate the feasibility to recognize silent speech by detecting the TMJ movement using active acoustic sensing.
- We propose a silent speech interface on commodity headphones that could be achieved with minimal or no hardware modifications.
- We evaluated HPSpeech on multiple commodity headphones with music playing through the headphones and demonstrated consistent performance.

2 METHODS

While speaking, either with or without sound, the mandible (the bone that forms the jaw) moves driven by facial muscles. Such movement can be detected at any position on the mandible, including the upper point - temporomandibular joint (TMJ) which is close to the ear, as illustrated in Figure 1(a). HPSpeech employs active acoustic sensing using the built-in speakers on commodity headphones to sense its movements. Sound waves are emitted by the speakers towards the ear and its surrounding areas, including the TMJ. Movements at the TMJ causes deformations on the surface of the skin, thus resulting in subtle changes in the sound traveling paths. Such changes can be captured by microphones inside the headphones and analyzed with methods such as echo profile analysis [22]. With this method, different utterances appear as distinct patterns on the echo profiles, as illustrated in Figure 1(c).

2.1 Data Processing and Deep Learning

HPSpeech utilizes active acoustic sensing as the sensing method, following a similar scheme as EarIO [22]. Specifically, HPSpeech employs 20-24kHz frequency-modulated continuous wave (FMCW) signals. To simulate the case when the user is using the headphones to play music, the inaudible signals are mixed into normal music and played through the built-in headphone speakers simultaneously.

After collecting the echoes from the microphones, we perform echo profile analysis following the scheme of EarIO [22] to obtain the echo profiles as the representation of the TMJ movements. Figure 1(c) presents sample echo profiles we collected for different silent speech commands.

We then use a customized deep learning pipeline to infer silent speech. The model contains a ResNet-18 backbone followed by a fully-connected decoder with Cross-Entropy Loss. An Adam optimizer is used with initial learning rate of 0.0002. The model was trained for 100 epochs using a single NVIDIA RTX 2080 Ti GPU. The batch size was set to 5. The data was collected first during the study and then the evaluation was conducted offline.

2.2 Microphone Positions

In order to examine the impact of microphone position, we conducted a pilot study with three researchers. We purchased a pair of headphones from the Internet (Razer Kraken [11]), removed the foam and then installed 8 microphones on the left side that evenly cover the internal area, as illustrated in Figure 2(a), and then

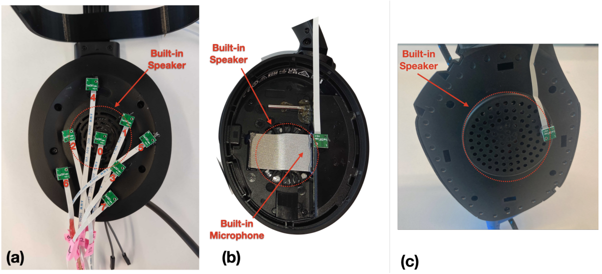

Figure 2: Headphones used and microphone positions tested. Only showing the left side. (a) Razer Kraken. 8 microphones positions were tested. (b) Bose QC 45. The installed microphone was close to the built-in microphone. (c) Xtream G1. It does not have built-in microphones inside.

installed the foam back. For the purpose of comparing different microphone positions, we only used the left side as the mandible is usually symmetric while speaking.

Following the same procedures as the main study, we evaluated the performance of each microphone position. Results show that all positions yielded very similar level of performance, ranging from 93.0% to 94.2%, which indicates that any microphone inside the headphone could produce satisfactory results. This is especially promising considering that commodity headphones with active noise cancellation usually have multiple built-in microphone inside.

2.3 Hardware Implementation

We installed a miniature microphone inside the headphones on each side, as illustrated in Figure 1(b). We then explored applying HPSpeech on different sets of headphones. We tested two sets of headphones during the study as illustrated in Figure 2(b-c): Bose QC 45 [4], which is one of the most popular noise-cancelling headphones, and Adesso Xtream G1 [10], a pair of gaming headphones, to further demonstrate that HPSpeech can be deployed on other headphones to provide an SSI.

As indicated in Section 2.2, any position would yield similar level of performance. We placed the microphone close to the location of the built-in microphone on the Bose QC 45 headphones, as illustrated in Figure 2(b). We chose this similar position to ensure that adopting HPSpeech on commodity headphones is highly feasible. We could not directly access the built-in microphone of Bose QC 45 due to lack of available APIs.

The built-in speakers and installed microphones were connected to a Teensy 4.1 micro-controller, which controlled the transmission and collection of sound waves. The data was first saved to an onboard micro SD card and then analyzed offline.

3 EVALUATION

We evaluated HPSpeech with a user study of 18 participants in total on two slightly modified commodity headphones approved by Cornell's institutional review board (IRB). For the first part, Bose QC 45 [4] was used and 10 participants (1 male, 9 female, average age 21.7, std 1.8) were recruited. For the second part, Adesso Xtream G1 [10] was used with 8 participants (2 male, 6 female, average age 27.0, std 9.5).

3.1 Commandset

To explore the practical use cases of HPSpeech, we are particularly interested in examining how HPSpeech can be used to control music playing through the headphones. We designed a set of 8 commands for controlling the music player using silent speech: **Previous, Next, Pause, Resume, Stop, Volume up, Volume down, Play**. These 8 commands cover common needs such as switching songs, controlling playback and pausing, as well as adjusting the volume.

3.2 Procedures

During the study, participants were instructed to silently utter the commands. As people tend to speak with smaller mouth movements when speaking silently, we instructed participants to speak with exaggerated mouth movements. The level of exaggeration did not cause visually "abnormal" speech behavior and is presented in the accompanying video. The study was split into 21 sessions, each lasting around 2 minutes. Participants were asked to utter each command 8 times in random order in each session and remount the device between sessions. To simulate real-world scenarios where music is playing while the user attempts to utter silent commands to control it, we selected 21 songs of different genres. During each session, a random song was mixed with our inaudible signals and played through the headphones. Active noise cancellation was not turned on during the study. As suggested by [9] we set the sound pressure level of our signal to less than 75dB.

After the data were collected, a researcher manually removed utterances where participants made a mistake or did not finish the command, which take up 0.86% of all utterances.

3.3 Results

We trained a user-dependent model for each participant. We treated the first session as practice and used the remainder 20 sessions for training/testing. We performed 10-fold cross-validation on these 20 sessions, using 18 sessions for training and 2 for testing each time. We summarize the performance of all participants in Figure 3(a). Results showed that the average accuracy with Bose QC 45 was 90.3%, std = 8.8%, with confusion matrix in Figure 3(d). Participant P09 had the worst performance of 67.7%, while performances on other participants were all over 85%. We examined the case of P09 and noticed that she did not move her mouth much during the study. Specifically, the jaw movements were very small, thus making it difficult to capture them with our system. We acknowledge that this is a limitation of the system.

HPSpeech used both sides of the headphones. Due to issues such as asymmetric head shape, head movements, etc., using both sides provided significantly better performance. Using only data on one side degraded the performance to around 83%.

The performance on Xtream G1 was more consistent, averaging 91.6% on 8 participants (std = 4.1%), as illustrated in Figure 3(b). We performed a one-way ANOVA test on the results from the first study and the second study. Results did not indicate significant performance difference (F(1, 16) = 0.146, p = 0.71 > 0.05) between

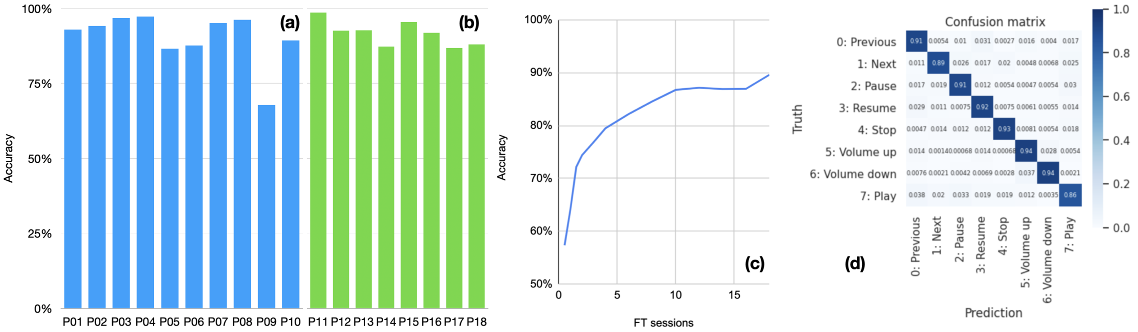

Figure 3: Evaluation results. (a) Performance on the 10 participants from the first study with Bose QC 45. (b) Performance on the 8 participants from the second study with Xtream G1. (c) Performance curve when different amount of training data from the same user is applied. (d) Confusion matrix.

the two headphones, indicating that HPSpeech can easily adapt to different headphones.

3.4 Training Effort

We would like to understand the impact of training length from each user in order to obtain satisfactory performance. We first trained a leave-one-participant-out (LOPO) model for each user. This was a user independent model, where new participants did not need to provide any training data. The average performance of this model was 46.9%, a lot better than random guess but far from satisfactory. This indicates that HPSpeech is still a user-dependent system. We then fine-tuned the LOPO model with different amounts of data from the same user and generated a performance curve in Figure 3(c). Results indicate that with only 4 sessions (8 minutes) of training data from the same user, HPSpeech can already achieve 80.0% accuracy. We discuss possible ways to further reduce training effort for deployment at scale in Section 4.

3.5 Comfort

After the study, we asked participants to rate the comfort level of the device from 0 (most uncomfortable) to 5 (most comfortable). The average comfort rating is 4.47 (std = 0.48), indicating that HPSpeech is very comfortable to wear during the study. Participants did mention that they could still hear some sounds in addition to the music. To address this issue, it is possible to further reduce the speaker power, or use sensing methods that do not involve sudden frequency change to reduce or remove frequency leakage.

4 DISCUSSION

We designed HPSpeech with the hope that it can be quickly deployed at scale. HPSpeech utilizes the existing speakers on the headphones, which significantly reduces hardware modifications. Theoretically, HPSpeech can also utilize the existing microphones inside the headphones, which are widely available on commodity noise-cancelling headphones. Unfortunately, we could not find open APIs that grant access to these microphones.

The system needs to be readily available with minimal amount of training effort from new users. In Section 3.4, we demonstrate that HPSpeech can reach 80% accuracy with only 8 minutes of training data from new users. This effort can be further reduced. For instance, EchoSpeech demonstrates that pre-training the model with other people's data can improve performance when the same amount of fine-tuning data is applied [33]. With this approach, it is possible to first collect a large amount of training data from other users and pre-train a large base model to further reduce the training effort from new users or eventually obtain a user-independent system.

Currently, HPSpeech needs to perform calculations offline. With current technology, it is possible to deploy the data processing pipelines on a smart phone, as demonstrated in EchoSpeech [33]. With advancement in embedded AI chips, it is even possible to perform the calculations inside the headphones in the future.

Of course, there are still issues that remain to be addressed before actually putting HPSpeech to deployment. For instance, due to lack of access to the built-in microphones, we could not fully implement the system on noise cancelling headphones. In addition, we did not test HPSpeech with noise cancellation turned on. We experienced strong interference between the noise cancellation and our system in our initial trials. It might be caused by incompatible sampling rate (HPSpeech uses 50kHz while common microphones usually use 44.1kHz or 48kHz) which causes the noise cancellation microphones to sample false low-frequency components. We believe that the noise cancellation system might need to be aware of our signal for it to work properly. A direct integration with the headphones' hardware is needed to thoroughly investigate and resolve this issue.

With 90% accuracy, it still means that HPSpeech makes 1 mistake every about 10 utterances, which could lead to user dissatisfaction. Further improving the performance could lead to better user experiences. Other issues such as the slight audible noise mentioned in Section 3.5 also needs to be addressed. In addition, one of the fundamental limitation is that HPSpeech requires users to speak with slightly exaggerated mouth movements in order to achieve reliable performance. We leave these issues for future exploration.

ACKNOWLEDGMENTS

The authors would like to thank all participants, reviewers and Cornell Information Science for this project. This material is based upon work supported by the National Science Foundation under Grant No. 2239569.

REFERENCES

[1] Abdelkareem Bedri, Himanshu Sahni, Pavleen Thukral, Thad Starner, David Byrd, Peter Presti, Gabriel Reyes, Maysam Ghovanloo, and Zehua Guo. 2015. Toward Silent-Speech Control of Consumer Wearables. *Computer* 48, 10 (2015), 54–62. https://doi.org/10.1109/MC.2015.310

[2] Lam A. Cheah., James M. Gilbert., Jose A. Gonzalez., Phil D. Green., Stephen R. Ell., Roger K. Moore., and Ed Holdsworth. 2018. A Wearable Silent Speech Interface based on Magnetic Sensors with Motion-Artefact Removal. In *Proceedings of the 11th International Joint Conference on Biomedical Engineering Systems and Technologies - BIODEVICES,*. INSTICC, SciTePress, 56–62. https://doi.org/10.5220/0006573200560062

[3] Tuochao Chen, Benjamin Steeper, Kinan Alsheikh, Songyun Tao, François Guimbretière, and Cheng Zhang. 2020. C-Face: Continuously Reconstructing Facial Expressions by Deep Learning Contours of the Face with Ear-Mounted Miniature Cameras. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology* (Virtual Event, USA) *(UIST '20)*. Association for Computing Machinery, New York, NY, USA, 112–125. https://doi.org/10.1145/3379337.3415879

[4] Bose Corporation. 2023. *QuietComfort 45 Noise Cancelling Smart Headphones | Bose.* Retrieved May 26, 2023 from https://www.bose.com/en_us/products/headphones/noise_cancelling_headphones/quietcomfort-headphones-45.html

[5] Tamás Gábor Csapó, Tamás Grósz, Gábor Gosztolya, László Tóth, and Alexandra Markó. 2017. DNN-Based Ultrasound-to-Speech Conversion for a Silent Speech Interface. In *Proc. Interspeech 2017*. 3672–3676. https://doi.org/10.21437/Interspeech.2017-939

[6] B. Denby, Y. Oussar, G. Dreyfus, and M. Stone. 2006. Prospects for a Silent Speech Interface using Ultrasound Imaging. In *2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings*, Vol. 1. I–I. https://doi.org/10.1109/ICASSP.2006.1660033

[7] Jose A. Gonzalez, Lam A. Cheah, Angel M. Gomez, Phil D. Green, James M. Gilbert, Stephen R. Ell, Roger K. Moore, and Ed Holdsworth. 2017. Direct Speech Reconstruction From Articulatory Sensor Data by Machine Learning. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 25, 12 (2017), 2362–2374. https://doi.org/10.1109/TASLP.2017.2757263

[8] Robin Hofe, Stephen R. Ell, Michael J. Fagan, James M. Gilbert, Phil D. Green, Roger K. Moore, and Sergey I. Rybchenko. 2013. Small-vocabulary speech recognition using a silent speech interface based on magnetic sensing. *Speech Communication* 55, 1 (2013), 22–32. https://doi.org/10.1016/j.specom.2012.02.001

[9] Carl Q Howard, Colin H Hansen, and Anthony C Zander. 2005. A review of current ultrasound exposure limits. *The Journal of Occupational Health and Safety of Australia and New Zealand* 21, 3 (2005), 253–257.

[10] Adesso Inc. 2023. *Adesso Xtream G1 Gaming Headphones.* Retrieved May 26, 2023 from https://www.adesso.com/product/xtream-g1-multimedia-gaming-headphone-headset-with-microphone/

[11] Adesso Inc. 2023. *Competitive Gaming Headset - Razer Kraken.* Retrieved May 26, 2023 from https://www.razer.com/gaming-headsets/razer-kraken

[12] Yincheng Jin, Yang Gao, Xuhai Xu, Seokmin Choi, Jiyang Li, Feng Liu, Zhengxiong Li, and Zhanpeng Jin. 2022. EarCommand: "Hearing" Your Silent Speech Commands In Ear. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 6, 2, Article 57 (jul 2022), 28 pages. https://doi.org/10.1145/3534613

[13] Eloi Moliner Juanpere and Tamás Gábor Csapó. 2019. Ultrasound-Based Silent Speech Interface Using Convolutional and Recurrent Neural Networks. *Acta Acustica united with Acustica* 105, 4 (2019), 587–590.

[14] Arnav Kapur, Shreyas Kapur, and Pattie Maes. 2018. AlterEgo: A Personalized Wearable Silent Speech Interface. In *23rd International Conference on Intelligent User Interfaces* (Tokyo, Japan) *(IUI '18)*. Association for Computing Machinery, New York, NY, USA, 43–53. https://doi.org/10.1145/3172944.3172977

[15] Arnav Kapur, Utkarsh Sarawgi, Eric Wadkins, Matthew Wu, Nora Hollenstein, and Pattie Maes. 2020. Non-Invasive Silent Speech Recognition in Multiple Sclerosis with Dysphonia. In *Proceedings of the Machine Learning for Health NeurIPS Workshop (Proceedings of Machine Learning Research, Vol. 116)*, Adrian V. Dalca, Matthew B.A. McDermott, Emily Alsentzer, Samuel G. Finlayson, Michael Oberst, Fabian Falck, and Brett Beaulieu-Jones (Eds.). PMLR, 25–38. https://proceedings.mlr.press/v116/kapur20a.html

[16] Myungjong Kim, Beiming Cao, Ted Mau, and Jun Wang. 2017. Speaker-Independent Silent Speech Recognition From Flesh-Point Articulatory Movements Using an LSTM Neural Network. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 25, 12 (2017), 2323–2336. https://doi.org/10.1109/TASLP.2017.2758999

[17] Myungjong Kim, Nordine Sebkhi, Beiming Cao, Maysam Ghovanloo, and Jun Wang. 2018. Preliminary Test of a Wireless Magnetic Tongue Tracking System for Silent Speech Interface. In *2018 IEEE Biomedical Circuits and Systems Conference (BioCAS)*. 1–4. https://doi.org/10.1109/BIOCAS.2018.8584786

[18] Naoki Kimura, Tan Gemicioglu, Jonathan Womack, Richard Li, Yuhui Zhao, Abdelkareem Bedri, Alex Olwal, Jun Rekimoto, and Thad Starner. 2021. Mobile, Hands-Free, Silent Speech Texting Using SilentSpeller. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI EA '21)*. Association for Computing Machinery, New York, NY, USA, Article 178, 5 pages. https://doi.org/10.1145/3411763.3451552

[19] Naoki Kimura, Tan Gemicioglu, Jonathan Womack, Richard Li, Yuhui Zhao, Abdelkareem Bedri, Zixiong Su, Alex Olwal, Jun Rekimoto, and Thad Starner. 2022. SilentSpeller: Towards Mobile, Hands-Free, Silent Speech Text Entry Using Electropalatography. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 288, 19 pages. https://doi.org/10.1145/3491102.3502015

[20] Naoki Kimura, Kentaro Hayashi, and Jun Rekimoto. 2020. TieLent: A Casual Neck-Mounted Mouth Capturing Device for Silent Speech Interaction. In *Proceedings of the International Conference on Advanced Visual Interfaces* (Salerno, Italy) *(AVI '20)*. Association for Computing Machinery, New York, NY, USA, Article 33, 8 pages. https://doi.org/10.1145/3399715.3399852

[21] Naoki Kimura, Michinari Kono, and Jun Rekimoto. 2019. SottoVoce: An Ultrasound Imaging-Based Silent Speech Interaction Using Deep Neural Networks. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–11. https://doi.org/10.1145/3290605.3300376

[22] Ke Li, Ruidong Zhang, Bo Liang, François Guimbretière, and Cheng Zhang. 2022. EarIO: A Low-Power Acoustic Sensing Earable for Continuously Tracking Detailed Facial Movements. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 6, 2, Article 62 (jul 2022), 24 pages. https://doi.org/10.1145/3534621

[23] Richard Li, Jason Wu, and Thad Starner. 2019. TongueBoard: An Oral Interface for Subtle Input. In *Proceedings of the 10th Augmented Human International Conference 2019* (Reims, France) *(AH2019)*. Association for Computing Machinery, New York, NY, USA, Article 1, 9 pages. https://doi.org/10.1145/3311823.3311831

[24] Hiroyuki Manabe, Akira Hiraiwa, and Toshiaki Sugimura. 2003. "Unvoiced Speech Recognition Using EMG - Mime Speech Recognition". In *CHI '03 Extended Abstracts on Human Factors in Computing Systems* (Ft. Lauderdale, Florida, USA) *(CHI EA '03)*. Association for Computing Machinery, New York, NY, USA, 794–795. https://doi.org/10.1145/765891.765996

[25] Geoffrey S Meltzner, James T Heaton, Yunbin Deng, Gianluca De Luca, Serge H Roy, and Joshua C Kline. 2018. Development of sEMG sensors and algorithms for silent speech recognition. *Journal of Neural Engineering* 15, 4 (jun 2018), 046031. https://doi.org/10.1088/1741-2552/aac965

[26] Himanshu Sahni, Abdelkareem Bedri, Gabriel Reyes, Pavleen Thukral, Zehua Guo, Thad Starner, and Maysam Ghovanloo. 2014. The Tongue and Ear Interface: A Wearable System for Silent Speech Recognition. In *Proceedings of the 2014 ACM International Symposium on Wearable Computers* (Seattle, Washington) *(ISWC '14)*. Association for Computing Machinery, New York, NY, USA, 47–54. https://doi.org/10.1145/2634317.2634322

[27] Tanja Schultz. 2010. ICCHP Keynote: Recognizing Silent and Weak Speech Based on Electromyography. In *Computers Helping People with Special Needs*, Klaus Miesenberger, Joachim Klaus, Wolfgang Zagler, and Arthur Karshmer (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 595–604.

[28] Tanmay Srivastava, Prerna Khanna, Shijia Pan, Phuc Nguyen, and Shubham Jain. 2022. MuteIt: Jaw Motion Based Unvoiced Command Recognition Using Earable. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 6, 3, Article 140 (sep 2022), 26 pages. https://doi.org/10.1145/3550281

[29] László Tóth, Gábor Gosztolya, Tamás Grósz, Alexandra Markó, and Tamás Gábor Csapó. 2018. Multi-Task Learning of Speech Recognition and Speech Synthesis Parameters for Ultrasound-based Silent Speech Interfaces. In *Proc. Interspeech 2018*. 3172–3176. https://doi.org/10.21437/Interspeech.2018-1078

[30] You Wang, Ming Zhang, RuMeng Wu, Han Gao, Meng Yang, Zhiyuan Luo, and Guang Li. 2020. Silent Speech Decoding Using Spectrogram Features Based on Neuromuscular Activities. *Brain Sciences* 10, 7 (2020). https://doi.org/10.3390/brainsci10070442

[31] Kele Xu, Yuxiang Wu, and Zhifeng Gao. 2019. Ultrasound-Based Silent Speech Interface Using Sequential Convolutional Auto-Encoder. In *Proceedings of the 27th ACM International Conference on Multimedia* (Nice, France) *(MM '19)*. Association for Computing Machinery, New York, NY, USA, 2194–2195. https://doi.org/10.1145/3343031.3350596

[32] Ruidong Zhang, Mingyang Chen, Benjamin Steeper, Yaxuan Li, Zihan Yan, Yizhuo Chen, Songyun Tao, Tuochao Chen, Hyunchul Lim, and Cheng Zhang. 2022. SpeeChin: A Smart Necklace for Silent Speech Recognition. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 5, 4, Article 192 (dec 2022), 23 pages. https://doi.org/10.1145/3494987

[33] Ruidong Zhang, Ke Li, Yihong Hao, Yufan Wang, Zhengnan Lai, François Guimbretière, and Cheng Zhang. 2023. EchoSpeech: Continuous Silent Speech Recognition on Minimally-Obtrusive Eyewear Powered by Acoustic Sensing. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 852, 18 pages. https://doi.org/10.1145/3544548.3580801

[34] Yongzhao Zhang, Yi-Chao Chen, Haonan Wang, and Xingyu Jin. 2021. CELIP: Ultrasonic-Based Lip Reading with Channel Estimation Approach for Virtual Reality Systems. In *Adjunct Proceedings of the 2021 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2021*

ACM International Symposium on Wearable Computers (Virtual, USA) *(UbiComp '21)*. Association for Computing Machinery, New York, NY, USA, 580–585. https://doi.org/10.1145/3460418.3480163