

Silent Speech Recognition for BCI - A Review

You Wang¹, Ming Zhang¹, Ruifen Hu¹, Nan Li² and Guang Li^{1*}

¹State Key Laboratory of Industrial Control Technology, Institute of Cyber Systems and Control, Zhejiang University, China

²Department of Engineering, University of Cambridge, UK

*Corresponding author: Guang Li, State Key Laboratory of Industrial Control Technology, Institute of Cyber Systems and Control, Zhejiang University, Hangzhou 310027, Zhejiang, China



ARTICLE INFO

Received: 📅 April 16, 2020

Published: 📅 April 27, 2020

Citation: You Wang, Ming Zhang, Ruifen Hu, Nan Li, Guang Li. Silent Speech Recognition for BCI - A Review. Biomed J Sci & Tech Res 27(2)-2020. BJSTR. MS.ID.004478.

Keywords: Silent Speech Interface; Bio signal Activities; Articulator Motions

ABSTRACT

Silent Speech Interface (SSI) is just a novel member in Brain-Computer Interface (BCI) family, which relies on decoding the speech-related bio signal activities or articulator motions. It makes recognition or synthesis on data from articulators collected using a variety of sensors. Research on SSI is now an active inter-disciplinary field, combining neuroscience, computer science and engineering. This review presents current the advances and critical issues in the development of SSI. Relevant methods, practices and challenges are also included.

Abbreviations: SSI: Silent Speech Interface; BCI: Brain-Computer Interface; EMG: Electromyography; HMM: Hidden Markov Model; DTW: Dynamic Time Warping; LDA: Linear Discriminant Analysis

Introduction

Acoustic speech is the most common and comfortable communication method among humans. Starting from the brain, acoustic signals are produced by the synergistic work of vocal organs such as lung, larynx and throat. Bio signals and articulatory activities are along with these physiological processes and have turned out to be potential to interpret the speech production [1]. One of the typical methods is to make use of such speech-related bio signal or articulatory motion as an alternative way for speech recognition. Differing from conventional Brain-Computer Interface (BCI) which directly decoding cortical brain activity, the Silent Speech Interface (SSI) always uses the neuromuscular or articulator activities to indirectly trace back to the neural information [2,3]. As a relatively novel way of BCI, SSI has advantages in implementation

and application. Compared with conventional BCI, this technology requires fewer channel signals and the signal detection is much more convenient due to data usually recorded on muscle surface or via a non-contact way [4]. SSI has unique advantages in comparison with speech interface as well. It mostly only depends on the relevant electrophysiology activities and will not be interfered by ambient noises, so it works well in noisy condition. Secondly, as SSI does not need to be emitted voice, so privacy of communication can be guaranteed, and will not interact with surrounding people, especially in public area. Finally, SSI is an excellent choice to help people with speech disabilities (i.e., laryngectomy or dumb patients) or to language training [1,5,6]. Figure 1 shows the overall diagram of SSI, from data acquisition to the final applications.

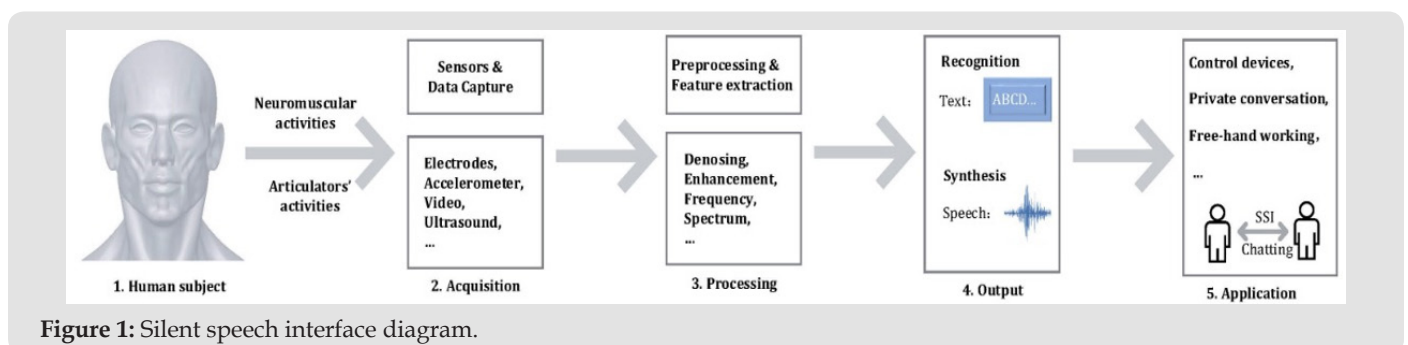


Figure 1: Silent speech interface diagram.

Data Acquisition

Some techniques have been employed to capture different speech-related bio signals, including articulatory and muscular activity. Articulatory movements (i.e., lip or tongue) usually come with vocal or whisper speech [7]. The popular one is relevant with neuromuscular activities, regardless of vocal or silent speech [5]. Measuring the motion of articulators, imaging and surface sensing are two common methods used in SSI. Video and ultrasound imaging can record the activities of visible or invisible speech articulators straightforward [8]. Because of its good safety and temporal resolution, imaging is well adopted to vocal tract and analysis under certain clinical conditions. Vibration or magnetic sensors used to capture the articulatory activity cause less signal crosstalk in multichannel systems and require less treatment on skin surface [7-9]. However, moving articulators do not work well in purely silent speech.

Both invasive and noninvasive ways have been used to record muscular activity, which is usually called Electromyography (EMG) [1]. Needle electrodes are inserted into muscle tissues and obtains high quality EMG with excellent spatial and temporal resolution invasively. So the medical expertise is required, and it is unsuitable for frequent use. Noninvasive EMG sensors are more popular for SSI. In accessible areas, always face and around, surface electrodes are placed on particular muscles or some designated grids to obtain surface EMG (sEMG) data [10-12]. Due to the skin and tissue between the electrodes and muscles, the signals measured are actually represented the signal mixed from several muscles. Although the signal quality is relatively poor, it is still a preferred implementation because of the convenience and hygiene in use.

Methods

There are only a few research on SSI based on the articulator motions. Ultrasonic and optical images are processed to obtain feature sets and then the speech information are recognized via silent vocoders [8,13-14]. Flexible and skin-attachable vibration sensor can perceive human voices quantitatively by the examined linear relationship between voice pressure and neck skin vibration [9]. For magnetic technique, permanent magnets attached to speech articulators measure the magnetic field for achieving speech recognition by Hidden Markov Model (HMM), Dynamic Time Warping (DTW) or other algorithms [15-17]. In EMG-based SSI, syllables, phonemes and spectrum are used to make recognitions under time, frequency and time-frequency domain. Initially, Linear Discriminant Analysis (LDA) and HMM are utilized as classifiers to make recognition [1,18]. Nowadays the approaches are more related to artificial neural networks that perform well in related studies [19].

Output and Potential Applications

SSI can output in two forms, silent speech recognition in text code and synthetic speech in voice [6]. It is up to the practical

requirements. Potentially SSI can be implement in following circumstances:

- a) Medical prostheses control and speech training by patients with speech disabilities
- b) Hands-free peripheral device control
- c) Communication in privacy or noisy ambience [20-24].

Open Challenges

Substantial progress has been achieved in recent years, especially in data acquisition approaches and algorithms. However, there are still some common challenges. Just as conventional speech recognition, large vocabulary datasets with high performance are required. Subject or speaker independence is another concern. The recognition is closely relevant with speaker's anatomy, so the differences of muscular movement of different individuals and sensor position of different trials may influence the accuracy of SSI. To practice in reality, conveniently wearable recording systems are required to work robustly.

Conclusion

In this paper, an overview of silent speech interface, a new proposed promising technology, is presented. Brief introduction to signal obtain approaches, recognition methods and challenges of SSI are illustrated. Currently, the accuracy of SSI is achieved more than 90% with reasonable sEMG data size. The noninvasive and convenient way is a promising method for BCI.

Acknowledgement

This work is supported by Zhejiang University Education Foundation Global Partnership Fund.

Conflict of interest

The authors declare that they have no competing interests.

References

1. Schultz T, Wand M, Hueber T, Krusienski DJ, Herff C, et al. (2017) Biosignal-based spoken communication: A survey. *IEEE Audio, Speech, Lang Process* 25(12): 2257-2271.
2. Meltzner GS, Heaton JT, Deng Y, De Luca G, Roy SH, et al. (2018) Development of sEMG sensors and algorithms for silent speech recognition. *J Neural Eng* 15(4): 046031.
3. Anumanchipalli GK, Chartier J, Chang EF (2019) Speech synthesis from neural decoding of spoken sentences. *Nature* 568(7753): 493-498.
4. Zhang M, Wang Y, Wei Z, Yang M, Luo Z, et al. (2020) Inductive conformal prediction for silent speech recognition. *J. Neural Eng.*
5. Denby B, Schultz T, Honda K, Hueber T, Gilbert JM, et al. (2010) Silent speech interfaces. *Speech Commun* 52(4): 270-287.
6. Wand M, Janke M, Schultz T (2014) Tackling speaking mode varieties in EMG-based speech recognition. *IEEE Trans Biomed Eng* 61(10): 2515-2526.
7. Hofe R, Ell SR, Fagan MJ, Gilbert JM, Green PD, et al. (2013) Small-vocabulary speech recognition using a silent speech interface based on magnetic sensing. *Speech Commun* 55(1): 22-32.

8. Hueber T, Benaroya EL, Gérard C, Denby B, Gérard D, et al. (2010) Development of a silent speech interface driven by ultrasound and optical images of the tongue and lips. *Speech Commun* 52(4): 288-300.
9. Lee S, Kim J, Yun I, Bae GY, Kim D, et al. (2019) An ultrathin conformable vibration-responsive electronic skin for quantitative vocal recognition. *Nature Commun* 10(1): 1-11.
10. Hakonen M, Piitulainen H, Visala A (2015) Current state of digital signal processing in myoelectric interfaces and related applications. *Biomed. Signal Process* 18(4): 334-359.
11. Schultz T (2010) Modeling coarticulation in EMG-based continuous speech recognition. *Speech Commun* 52(4): 341-353.
12. Meltzner GS, Heaton JT, Deng Y, De LG, Roy, et al. (2017) Silent speech recognition as an alternative communication device for persons with laryngectomy. *IEEE Audio, Speech, Lang Process* 25(12): 2386-2398.
13. Ji Y, Liu L, Wang H, Liu Z, Niu Z, et al. (2018) Updating the Silent Speech Challenge benchmark with deep learning. *Speech Commun* 98: 42-50.
14. Hueber T, Bailly G (2016) Statistical conversion of silent articulation into audible speech using full-covariance hmm. *Comput. Speech Lang* 36: 274-293.
15. Gonzalez JA, Cheah LA, Gilbert JM, Bai J, Ell SR, et al. (2016) A silent speech system based on permanent magnet articulography and direct synthesis. *Comput Speech Lang* 39(9): 67-87.
16. Gilbert JM, Rybchenko SI, Hofe R, Ell SR, Fagan MJ, et al. (2010) Isolated word recognition of silent speech using magnetic implants and sensors. *Med Eng Phys* 32(10): 1189-1197.
17. Kim M, Cao B, Mau T, Wang J (2017) Speaker-independent silent speech recognition from flesh-point articulatory movements using an LSTM neural network. *IEEE Audio, Speech, Lang Process* 25(12): 2323-2336.
18. Chan AD, Englehart K, Hudgins B, Lovely DF (2001) Myo-electric signals to augment speech recognition. *Med Biol Eng Comput* 39(4): 500-504.
19. Kapur A, Kapur S, Maes P (2018) Alterego: A personalized wearable silent speech interface. In *23rd Int. Conf. Intell. User Interfaces* pp. 43-53.
20. Scheme E J, Hudgins B, Parker PA (2007) Myoelectric signal classification for phoneme-based speech recognition. *IEEE Trans Biomed Eng* 54(4): 694-699.
21. Denby B, Chen S, Zheng Y, Xu K, Yang Y, et al. (2017) Recent results in silent speech interfaces. *J Acoust Soc Am* 141(5): 3646-3646.
22. Janke M, Diener L (2017) EMG-to-speech: Direct generation of speech from facial electromyographic signals. *IEEE Audio, Speech, Lang Process* 25(12): 2375-2385.
23. Armas W, Mamun KA, T Chau (2014) Vocal frequency estimation and voicing state prediction with surface emg pattern recognition. *Speech Commun* 63: 15-26.
24. Munna K, Mosarrat J (2018) Classification of myoelectric signal for sub-vocal hindi phoneme speech recognition. *J Intell Fuzzy Syst* 35(5): 1-8.

ISSN: 2574-1241

DOI: 10.26717/BJSTR.2020.27.004478

Guang Li. Biomed J Sci & Tech Res



This work is licensed under Creative Commons Attribution 4.0 License

Submission Link: <https://biomedres.us/submit-manuscript.php>



Assets of Publishing with us

- Global archiving of articles
- Immediate, unrestricted online access
- Rigorous Peer Review Process
- Authors Retain Copyrights
- Unique DOI for all articles

<https://biomedres.us/>