# Developing A Hands-Free Human-Computer Interface That Senses Tooth/Jaw-Movements with Headphones

Krishna Jha[1], Prahas Gantaram[1] and Ethan Kantz[1#]

[1]Academies of Loudoun
[#]Advisor

## ABSTRACT

People with motor disabilities such as amputated limbs are often incapable of using computers traditionally. Instead, they rely on other ways of computer use relying on sense facial movements by sending the signals from the sensor to an interface, and finally translating the signals into computer commands. The goal of this research is to create a hands-free Human Computer Interface (HCI) that detects sounds near the ears from teeth and jaw movements (e.g., clicking or grinding), and translates it into computer commands. Our research aims to make computers accessible to a wider range of people, as it does not require one to use their hands. The novelty of the proposed HCI is how it gathers input, which it does by receiving signals strictly from headphones, making it inexpensive and user-friendly. Our goal is to create a proof-of principle setup to show the feasibility of this HCI. To achieve this, a pair of headphones were wired to an m-audio amplifier, and the sound settings on the computer were set so that both the speaker/microphone were using the m-audio amplifier. A script written with Octave software plotted the recorded signals from the amplifier (recorded with headphones). While the script was running, facial movements, such as a tooth click, or jaw opening performed by the user generated audio signals that were recorded by the computer. The script then mapped the frequency distribution of the recorded waves. Our work so far shows that headphones can indeed gather inputs of different facial movements.

## Introduction

Millions of people with motor and speech disabilities like Amyotrophic Lateral Sclerosis (ALS), Duchenne Muscular Dystrophy (DMD), amputated limbs, and Broca's Aphasia cannot use computers traditionally, as they cannot type or use speech recognition technology. For example, Amyotrophic Lateral Sclerosis (ALS) is a hereditary disorder that damages a person's ability to walk and speak (National Institute of Neurological Disorders and Stroke, 2020) and affects up to 15,000 Americans (Centers for Disease Control and Prevention, 2017). Some of its most notable symptoms include muscle weakness and slurred speech (National Institute of Neurological Disorders and Stroke, 2020). Similarly, Duchenne Muscular Dystrophy (DMD) is another type of genetic disorder that weakens the arms, legs, lungs, heart, and throat (Centers for Disease Control and Prevention, 2020), which also prevents affected individuals from being able to move their limbs or speak. With an incidence of 14 per 100,000 people (Centers for Disease Control and Prevention, 2020), it is caused by mutations in the DMD gene that encodes dystrophin (U.S. National Library of Medicine, 2021). Aphasia is caused by stroke, brain infections, and gunshot wounds, affecting nearly a million Americans, with Broca's Aphasia being the most common type (National Institute on Deafness and Other Communication Disorders, 2017). Meanwhile, Broca's Aphasia is a disorder that results in paralysis of the arms and legs, as well as speech problems that prevent patients from being able to speak in long phrases (National Institute on Deafness and Other Communication Disorders, 2017). It also precludes comprehension of sentences with complex structure (National Aphasia Association, 2021).

Several types of hands-free Human-Computer Interfaces (HCIs) have been evaluated on patients with other motor disabilities, such as limb amputations. For example, Murphy et al. (2019) created a Brain-Computer Interface (BCI) system to help control prosthetic knees. The system was electroencephalogram (EEG) based, meaning that a non-invasive headset with electrodes would read electrical activity around the brain for a computer to interpret. It was found that the system's success rate was between 50-100%, which prompted the researchers to conclude that it was a workable BCI (Murphy et al., 2019). Similarly, Cohen et al. (2017) used functional magnetic resonance imaging (fMRI) to map brain activity as part of a Brain-Computer Interface (BCI) for patients to move prosthetic limbs. This would result in the user controlling a virtual avatar with a brain machine learning system that would analyze and convert the real time fMRI. It was found that amputees had about 92% control over the avatar, allowing the researchers to conclude that they made a possible BCI for amputees (Cohen et al., 2017).

There are many distinct types of hands-free HCI, and much research has been done on the topic. For example, Šumak et al. (2019) evaluated the effectiveness of an HCI utilizing the Emotiv EPOC+ neuroheadset headset, a type of headset that reads EEG signals, compared to existing methods of computer use by patients with ALS and DMD. When participants performed a series of tasks using the system, the efficiency of task completion among EPOC+ users were greater than that of users of other hands-free computer use methods (Šumak et al., 2019). Additionally, Vasanthan et al. (2012) used a webcam and Python-based algorithm to measure Region of Interest (ROI) coordinates, which are thresholds of facial movement, on a person's face. In this system, when a part of the face moved inside or outside of an ROI coordinate, the camera would detect this, allowing for the user input to be translated into computer commands. For example, if the user opened his or her eye, there would be specific ROI coordinates that would track the location of the eyelids and utilize that as user input. This input was processed and translated into commands used to control a two-wheeled Arduino robot. In general, the participants were able to control the robot with high (95%) accuracy (Vasanthan et al., 2012).

The research presented in this study is an extension of a proof of concept that tooth activity can be sensed by headphones. This is because Prakash et al. (2020) used an earphone interface to identify teeth gestures, such as tooth clicks and tooth grinds. Teeth activity can be sensed with just the headphones because vibrations from the jaw (and from the teeth by extension) travel across the skull, eventually reaching the ear where the headphone can sense them, though they were not amplified with an amplifier. However, they were processed with continuous wavelet transforms (CWT) because this type of wavelet analysis is better at mapping random spikes within a wave rather than a Fourier transform that analyzes the frequency content of a sinusoidal wave. This interface could accurately identify tooth gestures (from a set of seven different gestures) 90% of the time. The HCI proposed in this paper can sense tooth movements as well as other kinds of facial movements, like jaw openings and tongue movements, adding to its versatility.

For this project, signal processing in the form of wavelet transforms is necessary to analyze differences between the waveforms generated by different facial movements. To do this, it is important to recognize the most effective forms of signal processing as it relates to our data, as the signal processing technique used is heavily dependent on the form of the data. For example, the Mallat signal processing algorithm is mostly used for image processing while the continuous wavelet transform is primarily used on wavelet spikes (Shensa, 1992). Since the HCI was designed such that the signal data is in waveforms, the continuous wavelet transform is the most applicable form of signal and data processing that can be used in this study.

The purpose of this research is to create a device that can sense facial movements through headphones and translate them into computer commands. Since this study focuses more on the proof of concept of the sensing principle of the HCI (headphone sensing), it is investigating the effect of a user's facial movements, especially teeth/jaw clicking or grinding, on the signals sensed by the headphones. The constants in this research include the environment of the lab in terms of white noise, the type of headphone used, and the signal processing software used. We hypothesized that if the user performs a certain facial movement (a jaw or tooth click, for example), then the waveforms of the signals collected by the headphones for that facial movement will spike. This is because the facial movements affect the musculoskeletal system around the ear, with the headphones sensing this disturbance and using it as user input for a computer command. Therefore, the independent variable would be the facial movement, while the dependent variable would be the headphone readings. This means that the control should be the situation when the user has a neutral expression. For our setup, a pair of headphones were wired to an amplifier (the m-audio m track 8), which was connected to a computer. Then, signal processing via the Octave software was used to plot the wavelet transform of the recorded signal data. The significance of our research is that with other types of HCIs, this enables patients with motor disabilities to use more computer commands, thereby making hands-free computer use much easier for these patients.

## Materials

### Software

Before the headphones, amplifier, and computer were set up, the GNU Octave software (version 6.3.0), along with a few important packages, were installed. The first of these was the Large Time-Frequency Analysis Toolbox (LTFAT) package, which aided in conducting wavelet transforms on the headphone signal data as well as plotting these wavelet transforms. The "audio" package was also installed to allow for audio recording within the Octave software. Both packages were installed in the Octave command window by using the command "package install -forge packagename", where "packagename" represents the name of the Octave package. After installation, a script was written in Octave that allowed recording and wavelet transformation of headphone signals. This script is displayed in Figure 1 below.

```
1 pkg load ltfat;
2 pkg load audio;
3 b = record(5,44100);
4 [c,info] = fwt(b,'db8',10);
5 subplot(221);
6   plot(b);
7 subplot(222);
8   plotwavelets(c,info,44100,'dynrange',90);
```

**Figure 1.** GNU Octave Code.

In the first two lines, the two packages that were previously installed were loaded. The record() function in the third line records the signals coming from the amplifier (and from the headphones) for five seconds at a sampling rate of 44100 samples per second. In line four, a fast wavelet transform was performed on the recorded audio signals (assigned to the matrix variable 'b') with a db8 filter with 10 filter banks as recommended parameters (according to the Octave LTFAT documentation). Then, in lines five through eight, two subplots were created to plot the graphs of the signals. While the raw signal data was in subplot 221, the plotted wavelet transform was in subplot 222. The plotwavelets() command in line eight created the plotted wavelet transform based on the transform coefficients (the variables 'c' and 'info') as well as the sampling rate (44100).

### Hardware

The headphones used for measuring sound signals near the ear were simply Sony - ZX Series Wired On-Ear Headphones. The headphone jack was fitted with an amplifier plug (VCE 6.35mm (1/4 inch) Male to 3.5mm (1/8 inch) Female Stereo Audio Jack Adapter for Aux Cable, Guitar Amplifier, Headphone) that allowed for the headphone jack to be inserted into an input port (Port 1, for example) in the amplifier such that its signals could be amplified. The amplifier used in this experiment was the M-audio M Track 8 mixer, a type of mixer used in recording and making music. This amplifier had a charger associated with it (19V 100-240V DC 2.37A PIN POSITIVE 50-60HZ - M-AUDIO - SPSAFSP045RECN2) that was plugged into the power outlet to supply power to the mixer. Then, using a USB Type B cable, the amplifier was connected to a computer so that the computer could display the signals from the amplifier on Octave. The two figures below show the system architecture and the actual setup of the HCI, respectively.

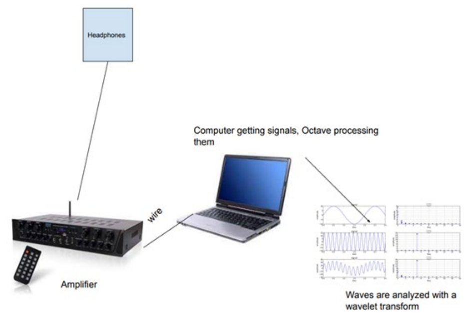


**Figure 2.** System Architecture of HCI.

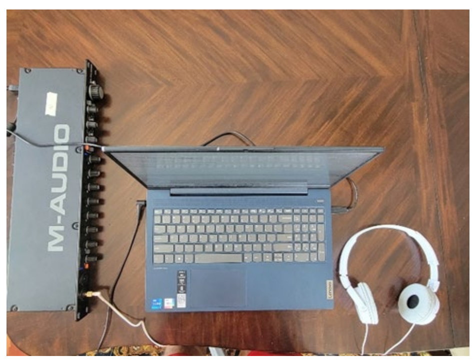


**Figure 3.** Picture of HCI Setup.

## Methods

After the hardware is set up, the sound settings of the computer were accessed and changed such that both the speaker and microphone of the computer were set to the m-audio mixer, allowing the computer to receive signals from the amplifier. Then, the user would put the headphones on such that the bottom part of the headphone was at least partly touching the part of the jaw surrounding the ear. After the Octave script was run by the user, he or she would perform a particular type of facial movement (such as a jaw opening, for example) during the five second recording duration. Once the five second duration concluded, two graphs would appear: one of the raw signal data of the facial movements of the user and the plotted wavelet transform of the raw signals. This process was repeated for jaw openings, tooth clicks, and tongue movements for several trials, though only one of each trial is shown in the data.

## Results

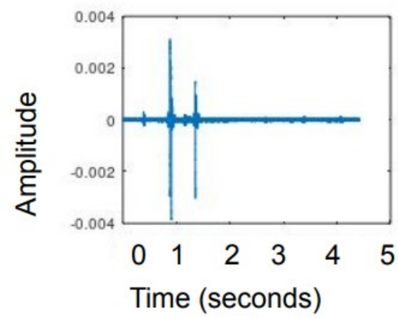


**Figure 4.** Signal Data for Two Tooth Click Movements. In Figures 4-6, each wavespike at an amplitude of around or above 0.002 shows that a facial movement has been performed during that time.

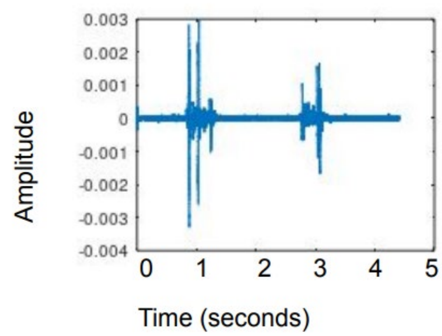


**Figure 5.** Signal Data for Two Jaw Openings

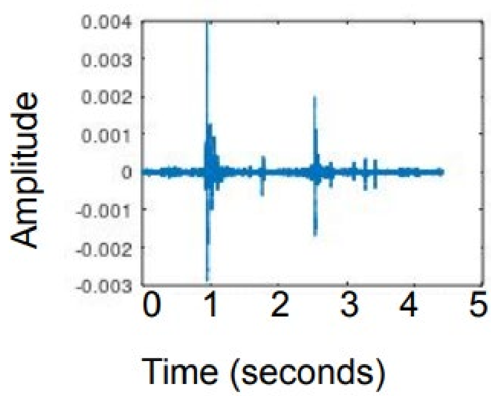


**Figure 6.** Signal Data for Two Tongue Sweep Movements

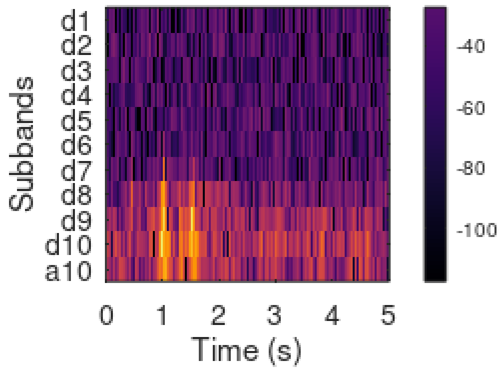


**Figure 7.** Wavelet Transformed Graph for Two Tooth Click Movements. The wavelet transformed graphs (Figures 7-9) each describe energy frequency: sections with more yellow are areas with higher frequency waves. These frequencies are found because this is a result of a fast wavelet transform (depicted in Figure 1).

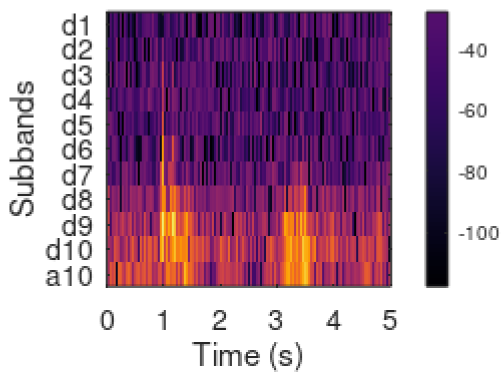


**Figure 8.** Wavelet Transformed Graph for Two Jaw Openings

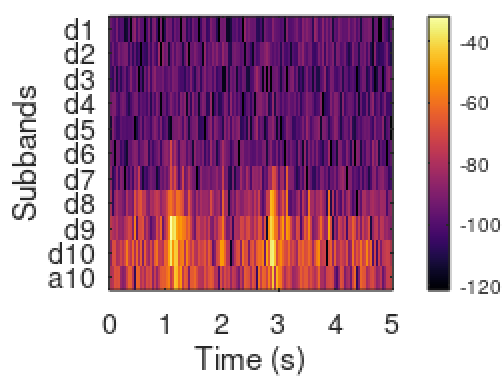


**Figure 9.** Wavelet Transformed Graph for Two Tongue Sweep Movements

# Discussion

Overall, the hypothesis that facial movements would cause changes within the waveform was supported by the data and results. As shown in Figure 4, when two tooth clicks were performed by the user over a five second span, two sharp spikes in the waveform occurred within the first two seconds (when the tooth clicks were performed). This also extended to Figure 7, where the two yellow spikes show a high energy frequency (or a spike) within the wavelet. This was likely the case because a tooth click can only create one sound (from the actual collision of the teeth themselves), resulting in only a minute spike showing up in the waveform. In Figure 5, the signals displayed from the user performing two jaw movements are noticeably different from those of the tooth clicks because the jaw movement causes a group of spikes to appear rather than just the one spike that occurs due to the tooth click. This phenomenon can also be seen in the wavelet transformed graph (Figure 8), though Figure 8 does seem to have regions of high signal energy that span for a longer period than those of Figure 7 (tooth clicks). This is likely due to the fact that a jaw movement simply lasts longer than tooth clicks. Jaw movements most likely generated these waveforms because the headphone not only senses the opening of the jaw, but also the closing of the jaw. Additionally, the jaw movement simply requires more exertion of muscles than a tooth click. As for the tongue sweep graph (Figure 6), a large spike is followed by a series of small spikes for each tongue sweep, which is likely the case because the tongue sweep was done so that the tongue started from the right side of the mouth from the user's perspective (at the top teeth) to the left side of the mouth. This means that the headphones sensed a stronger signal (the large spike) at the beginning, as the tongue had just started moving. Then, as the tongue traveled to the left side of the mouth from the user's reference frame, smaller spikes were detected as the tongue sweep concluded. Though all these data points suggest the functionality of the proof of concept of the HCI, more data is needed to make definitive conclusions.

Compared to earlier research on hands-free HCIs, the results in this study are similar in that facial movements result in differences in signal waveforms from the respective sensors used. However, many other researchers have measured the success rate of their hands-free HCIs in translating user input into computer commands. They would accomplish this by having the user perform a specific set of tasks, all of which required different facial movements from the HCI. Then, they would not only measure the success rate of the user completing his or her task, but also measure the rate at which the hands-free HCI converted facial movements into commands. However, this was not possible in this study because the device was only tested on the researchers themselves, who did not have any motor disabilities or impairments. This severely limited the ability to measure the practicality of the HCI presented in this study because it was not evaluated on users who fit the target population. This was the main reason that this research primarily focused on the proof of concept of the sensing mechanism of the HCI rather than how effective it was (its success rate) in allowing patients with motor disabilities to use computers. In this respect, the data from the HCI presented in this study suggests that this could possibly be a useful method of hands-free computer use for patients with motor disabilities, especially in limb amputees who have lost the ability to speak, but can move their jaw, tongue, and teeth. However, as stated earlier, more data is needed to solidify this conclusion.

In the future, it will be important to investigate correlations between facial movements sensed by the headphones and electrical signals, as myoelectric signals triggered by the brain to regulate facial muscle movement can be sensed and used to facilitate hands-free human to computer interaction. This would likely be done by building small myoelectric sensors that could be placed around or in the ear in a non-invasive manner. Then, as these myoelectric sensors are sensing electrical signals triggered by facial movements, the headphones are simultaneously on the user and are also recording signals. This would allow for correlations between different waveforms coming from headphones and myoelectric sensors to be investigated, further increasing the diversification of hands-free computer use for patients with motor disabilities as well as increased facial gesture recognition used to facilitate HCIs, as one HCI can be the backup for another one in case one of them fails.

## References

Centers for Disease Control and Prevention. (2017). Amyotrophic lateral sclerosis. *CDC*. https://www.cdc.gov/als/WhatisALS.html

Centers for Disease Control and Prevention. (2020). What is Muscular Dystrophy?. *CDC*. https://www.cdc.gov/ncbddd/musculardystrophy/facts.html

Cohen, O., Doron, D., Koppel, M., Malach, R., & Friedman, D. (2019). High Performance BCI in Controlling an Avatar Using the Missing Hand Representation in Long Term Amputees. *SpringerBriefs in Electrical and Computer Engineering*, 93-101. https://doi.org/10.1007/978-3-030-05668-1_9

Murphy, D. P., Bai, O., Gorgey, A. S., Fox, J., Lovegreen, W. T., Burkhardt, B. W., Atri, R., Marquez, J. S., Li Q., & Fei, D. (2017). Electroencephalogram-Based Brain–Computer Interface and Lower-Limb Prosthesis Control: A Case Study. *Frontiers in Neurology, 8*. https://doi.org/10.3389/fneur.2017.00696

National Aphasia Association. (2021, February 17). Aphasia Fact Sheet. *NAA*. Retrieved from https://www.aphasia.org/aphasia-resources/aphasia-factsheet/?gclid=CjwKCAiAmrOBBhA0EiwArn3mfGvkt5YjhR1aqyd_lumcFWkf7d6rvXr8xbO8bHuFODfwC1CRcGhPABoComcQAvD_BwE

National Institute on Deafness and Other Communication Disorders. (2017). Aphasia. *NIH*. https://www.nidcd.nih.gov/health/aphasia

Prakash, J., Yang, Z., Wei, Y. L., Hassanieh, H., Chaudhary, R. R. (2020). EarSense: earphones as a teeth activity sensor. Proceedings of the 26th Annual International Conference on Mobile Computing and Networking. https://doi.org/10.1145/3372224.3419197

Shensa, M.J. (1992). The Discrete Wavelet Transform: Wedding the A Trous and Mallat Algorithms. *IEEE Transactions on Signal Processing, 40(10)*, 2464–2482. https://doi.org/10.1109/78.157290