Deciphering Sounds Through Patterns of Vibration on the Skin

Michael V. Perrotta¹, Thorhildur Asgeirsdottir¹, David M. Eagleman¹,²*

¹Neosensory, San Mateo, CA, USA
²Department of Psychiatry and Behavioral Sciences, Stanford University, Stanford, CA, USA

*Correspondence:
David M. Eagleman, davideagleman@stanford.edu

Keywords: sensory substitution, haptics, sound identification, hearing loss, deafness.

Word count: 3200

Abstract
Sensory substitution refers to the concept of feeding information to the brain via an atypical sensory pathway. We here examined the degree to which participants (deaf and hard of hearing) can learn to identify sounds that are algorithmically translated into spatiotemporal patterns of vibration on the skin of the wrist. In a three-alternative forced choice task, participants could determine the identity of up to 94.6% of the stimuli simply by the spatial pattern of vibrations on the skin. Performance improved significantly over the course of 1 month. Younger participants tended to score better, possibly because of higher brain plasticity, more sensitive skin, or better skills at playing digital games. Similar results were obtained with pattern discrimination, in which a pattern representing the sound of one word was presented to the skin, followed by that of a second word. Participants answered whether the word was the same or different. With minimal difference pairs (distinguished by only one phoneme, such as “house” and “mouse”), the best performance was 83%, while with non-minimal pairs (such as house and zip) the best performance was 100%. Collectively, these results demonstrate that participants are capable of using the channel of the skin to interpret auditory stimuli, opening the way for low-cost, wearable sensory substitution for the deaf and hard of hearing communities.

1 Introduction
Sensory substitution, the method of passing sensory information through a different sensory pathway, has been documented for decades (Eagleman, 2020). Typically, this involves translating visual information from a video feed into touch on the skin (Bach-y-Rita et al., 1969), into touch on the tongue (Danilov and Tyler, 2005; Grant et al., 2016), or into a soundscape (Amedi et al., 2007; Auvray et al., 2007; Ward and Meijer, 2010; Striem-Amit et al., 2012). With such approaches, blind users can learn to identify visual objects and scenes.

Sensory substitution devices have also been created and tested for different modalities, including balance-to-touch to mitigate balance disorders (Tyler et al., 2003), sound-to-touch to aid people with hearing loss and deafness (Reed and Delhorne, 2003), and touch-to-touch to give somatosensation to someone with nerve damage (e.g. from leprosy; Bach-y-Rita, 1999) or who uses a prosthetic limb (Riso, 1999).
Such devices can help people with sensory impairments by letting them access the sensory information they are otherwise unable to. Although sensory substitution has been experimentally fruitful, it has suffered from lack of practicality due to size, expense, or inconvenience of wearing (Bach-y-Rita, 1983; Maidenbaum et al., 2014).

In that light, we have developed a practical, convenient device for sensory substitution: a wristband consisting of four vibratory motors, a microphone to capture sound, and a sophisticated processing system to convert audio in real time to spatiotemporal patterns of vibration; more in Methods below and in (Novich and Eagleman, 2015). With this approach, participants with deafness or hearing loss can learn to identify sound sources through the vibrations alone.

We here test the efficacy of this device as a sound-to-touch sensory substitution device. We set out to quantify whether such a device, worn on the wrist, and containing just four motors, would be able to convey enough information to a user that they could learn to differentiate similar patterns of vibration and identify sound sources from the vibrations.

Knowing that successful understanding of information from a sensory substitution device relies on neuroplastic changes (Rauschecker, 1995; Bach-y-Rita and Kercel, 2003; Bach-Y-Rita, 2004; Bubic et al., 2010; Proulx et al., 2014; Eagleman, 2020) and presuming that real-world use of the device should be useful here, we tested deaf and hard of hearing participants over the course of a month and had them take home the wristband during this time, allowing them to use it however they wished in their daily life.

We here analyze the performance of our participants, assessing whether users can identify sounds and differentiate similar patterns of vibration.

2 Methods

2.1 Participants

We tested 18 participants (eight male) ranging in age from 23 to 84 (median=59). No participant had any prior experience with the wristband. Three additional participants dropped out before the end of the study; their data is not included in this analysis. Demographic data of the 18 participants is in Table 1.

2.2 Wristband

2.2.1 Device

The Neosensory Buzz wristband consists of four vibratory motors built into the strap of the wristband (Figure 1a). The motors are linear resonant actuators (LRAs), vibrating at 175 Hz, capable of rising from 0 to 50% of their intensity within 30 ms. The motor intensity can be controlled with an eight bit resolution, meaning the motors can be controlled at 256 different intensities. At the highest intensity, each motor vibrates at 1.7 G_{RMS}. The motors are separated from one another at a distance of 18.2 mm and 19.2 mm for the small and large wristband sizes, respectively (center-to-center distances).

The top of the wristband is a module that contains the power button, user setting buttons, a microphone, and a microcontroller. The microphone captures audio in 16 ms chunks and sends each 16 ms chunk of audio to the microcontroller. The microcontroller processes the audio data through...
Neosensory’s sound-to-touch algorithm and vibrates the motors according to the output of the algorithm.

2.2.2 Algorithm

The algorithm is frequency-based, using a discrete Fourier transform to analyze the amplitudes of different frequencies present in the sound in bins from 300 to 7500 Hz (Figure 1b). The algorithm is updated every 16 ms. In the algorithm used here, one frequency is presented on the wristband in any frame – often this is the frequency with the highest amplitude, but not always. The algorithm implements several different filters to determine, in any given frame, which frequency (if any) should be represented on the wristband, and at what intensity. The intensity of the vibration follows Weber's Law (Geldard, 1957), which dictates that two high intensity stimuli require a greater difference in intensity than two low intensity stimuli for a participant to recognize a difference. Following this, intensities are activated on an exponential curve. The algorithm then translates this frequency and intensity into a motor output by mapping frequency onto one of 256 different spatial locations, accomplished by leveraging a well-known haptic illusion (Alles, 1970; Luzhnica et al., 2017). Specifically, an illusory location is a point on the wrist that is felt by the wearer of the wristband even when a motor is not located directly at that point. We stimulate these illusion locations by turning on two motors, one on either side of the illusion location, at specific intensities such that the wearer feels as if a single point somewhere between the two motors is vibrating.

2.3 Tests

2.3.1 Sessions

Each participant came into our laboratory three times: on day 0 (the day they received their wristband), day 14, and day 28. During each session participants completed three tasks: Sound Identification task, Pattern Discrimination task, and Minimum Threshold task. Once over the course of the study, participants also completed a demographic questionnaire and a hearing test. Between sessions, participants took the wristband home and were instructed to wear the wristband as they liked, as long as they wore it for at least four hours each day.

2.3.2 Demographics

Participants completed a questionnaire about their sex, age, and hearing loss profile (e.g. When did you lose your hearing? How much of the time do you use ASL to communicate with friends and family?).

2.3.3 Audiogram

We measured participants’ hearing loss using the Mimi hearing test (apps.apple.com) and Beyerdynamic Aventho wired on-ear headphones. Data for each participant is in Table 1.
2.3.4 Sound identification

Participants performed a task to determine if they could identify sounds based on the patterns of vibrations on their wrist. In each trial, the wristband presented a pattern of vibrations that lasted between 2 and 5 sec. (No audible sound was played, only the vibrations). Then the participant chose from a list of three options (e.g., dog bark, siren, running water) the one they believed corresponded with the pattern. This list was visible to the participant as soon as they started to feel the vibrations. Once the participant answered, they were shown which answer was correct.

The patterns of vibration were created by passing audio recordings of different sounds through the sound-to-touch algorithm offline and capturing the motor outputs associated with each sound’s audio recording (Figure 2). Sounds were chosen from a list of 14 sound classes: baby crying, car horn, car passing, clapping, clock alarm, coughing, dog bark, door knock, laughing, ringtone, running water, siren, smoke alarm, and speech. Each of the 14 sound classes had five sound recordings associated with it for a total of 70 audio recordings, each of which was translated into vibrational patterns. In each trial, one of the five patterns for the target sound source was chosen, such that the participant would not reencounter the exact same vibration pattern but instead was presented with similar but different vibration patterns for a given sound class. The 14 sounds were evenly distributed across trials.

Participants completed 56 trials of the Sound Identification task in each of the three sessions.

2.3.5 Pattern discrimination

Participants performed a task to determine whether they could distinguish different patterns of vibration on their wrist. In each trial, the participant was presented with two vibration patterns of ~1 sec, with 300-600 ms of silence between vibration patterns. The participant then answered whether the two vibration patterns were identical or different. Feedback told the participant whether an answer was correct or incorrect.

The vibration patterns were created by passing audio recordings of spoken words through the sound-to-touch algorithm offline and capturing the motor outputs associated with each word’s audio recording (Figure 3).

For half of the trials, the two vibration patterns were identical (i.e. created from the same audio recording of a single word). For the other half, the two vibration patterns were created from audio recordings of two different words. Of these trials, half were “minimal pairs” and half were “non-minimal pairs”. In a “minimal pair” trial, the two words used to create the vibration patterns differed from each other by only one phoneme (e.g. bees/cheese, rip/zip, bad/bed). For each minimal pair trial, one of 262 minimal pairs was randomly chosen. In non-minimal pair trials, the two words used to create the vibration patterns were chosen randomly from the 524 words that made up the 262 minimal pairs (e.g. bees/zip or rip/bed).

Participants completed 60 trials of the Pattern Discrimination task in each of the three sessions.

2.3.6 Minimum threshold

Participants performed a task which determined the lowest detectable intensity of vibration on their skin. In this task, participants felt pulses of vibrations at a single location on their wrist. The pulses were each 32 ms long and separated by 300 ms. The participant was instructed to press and hold a
button as soon as s/he could feel any vibration, and to let go of the button as soon as the pulsing vibrations became too weak to feel. When the button was pressed, the vibrations became less intense; when the button was not pressed, the vibrations became more intense. Figure 4 shows how the vibration intensities change as a participant indicates s/he can or cannot feel the vibrations. Each time the participant pressed or released the button, the amount of change between each intensity decreased, giving the opportunity to more precisely indicate when vibrations were being felt. Once the participant pressed and released the button six times, the trial ended.

The minimum threshold for a given trial was calculated as the average of the intensities at which the participant pressed and released the button. These averages exclude the first three presses and releases and only include the last three presses and releases (as indicated by the red pluses in Figure 4), since the rate of change of the intensities is lowest during the end of the trial and thus this reduces the error introduced by a participant being slow to react to their sensation.

We quantified the minimum threshold for five locations on each participant. Three of the locations were individual motors on the wristband while the other two were “illusion locations”, created by vibrating two motors simultaneously (see Algorithm section above). A participant’s final minimum threshold was calculated as the average of their minimum threshold at all five locations.

2.3.7 Institutional Review Board

The study protocol was approved by Solutions IRB. All subjects gave written informed consent in accordance with the Declaration of Helsinki.

3 Results

3.1 Sound identification

Remarkably, in their first session, 11 out of 18 of the participants were able to identify more than 62% of sounds simply by feeling these sounds as vibration patterns. The highest score a participant received in any of the three sessions was 94.6% (chance performance was 33.3%) (Figure 5A).

Over the course of one month there was a significant improvement. A one-way Anova test reveals a significant difference due to session day (F(2, 34) = 10.29, p < .001) and a post hoc test shows significantly higher scores at each session day; a significant improvement in score could be detected in these 18 subjects even after only 14 days of wearing the wristband. Day 0 to 14: r(17) = -.208, p = .027; day 0 to 28: r(17) = -.22, p < .001; day 14 to 28: r(17) = -.267, p = .016; tests are one-tailed paired-samples t-tests and p-values are adjusted with the Holm-Bonferroni method. The average score on day 0 was 65.1% while the average score on day 28 was 74.5%.

Finally, we saw a significant correlation when comparing participants’ average sound identification score to participants’ age (r(16) = -.46, p = .028; one-tailed; Figure 5B). This correlation suggests older participants perform worse on the sound identification task. Additionally, when investigating just the day 28 scores rather than the average of the three sessions, we find an even stronger correlation between score and age (r(16) = -.60, p = .005).

3.2 Pattern discrimination

We found participants were able to discriminate between similar vibrational patterns played through the wristband, with one participant scoring an average of 87.8% on non-minimal pair trials (Figure...
A). Being a two-alternative forced choice task, participants would be expected to score at 50% if they could not discriminate between patterns (i.e. chance). Participants performed significantly above chance for both the minimal and non-minimal pairs conditions (one-tailed, single-sample Student’s t-test: non-minimal pairs t(17) = 8.61, p < 0.001; minimal-pairs t(17) = 6.69, p < 0.001). Participants scored significantly higher on non-minimal pair trials than they did on minimal-pair trials (t(17) = 3.78, p < 0.001). We did not measure a statistically significant improvement in score from the first session to the third; we therefore averaged the data over the three sessions (Day 0, 14, 28). Further studies will determine whether a longer term of use improves discrimination.

Similar to the sound identification task, there was a correlation between age and performance on the pattern discrimination task. Although not quite significant, the trend suggests that participants who are older are more likely to score lower (r(16) = -0.31, p = .075; Figure 6B).

### 3.3 Minimum threshold

When taking the average minimum threshold score over the three sessions, we found participants tended to have a higher minimum threshold if they were older. This finding was not significant but was trending (r(16) = 0.29, p = .122; Figure 7).

### 4 Discussion

The purpose of this study was to investigate a pragmatic sound-to-touch sensory substitution device. Our results show that deaf and hard of hearing participants are able to identify sound sources by feeling vibrations on their wrists, and in some cases can score nearly perfectly in a three-alternative forced choice task of sound identification. Moreover, participants performed above chance on their first day, even before any experience wearing the wristband around sounds in their everyday lives. Performance then increased significantly over time.

On tests of pattern discrimination, participants could significantly detect differences between two words that differed by only a single phoneme (minimal pairs) and by more than a single phoneme (i.e. non-minimal pairs). Future studies will determine whether the resolution of the wristband (spatial and/or temporal) is simply too low to resolve single phoneme differences.

Across all tasks, older participants tended to have a lower performance. This was significant in the sound identification task, and strongly trending in the pattern discrimination task and the minimum threshold task. It is possible the trend might become significant with a greater number of participants.

There are several possibilities for the age-performance correlation. First, previous studies show significant decreases in vibration thresholds with age (Verrillo, 1980; Gescheider et al., 1996; Deshpande et al., 2008). It stands to reason that as the skin loses sensitivity a participant receives less haptic information (because they cannot feel the weak vibrations) and thus has more difficulty identifying a sound source that caused a pattern of vibrations or discriminating two similar patterns of vibration. A second reason may be that older participants are worse at learning due to decreased neuroplasticity (Kempermann et al., 2002; Lu et al., 2004). Although decreased learning would not by itself explain a correlation between age and score on day 0 (which we did see), it could explain why younger participants improved faster. Other possible reasons why older participants scored worse might be that they had lower attention (Wright and Elias, 1979; McDowd and Birren, 1990; Bolton and Staines, 2012), were less familiar with computer-based tests and games, or had worse two-point discrimination on the skin (Shimokata and Kuzuya, 1995; Bowden and McNulty, 2013).
5 Conclusion

We conclude that a practical, wrist-worn sensory substitution device with four vibrating motors is capable of providing enough information to a deaf or hard of hearing user that they are able to identify sound sources through patterns of vibrations.

Our next steps are to run studies that quantify performance over a longer period of time and with a larger number of participants. Studies over a matter of months will allow us to assess performance changes over time and to determine how high performance can get. Many studies of cochlear implants wait six months after implantation before running sound identification tasks and continue to see improvements over years following implantation (Grant et al., 1999; Reed and Delhorne, 2005; Shafiro et al., 2015; Strelnikov et al., 2018). This suggests we might continue to see improvement in performance with the wristband over the course of years.

Future studies will also investigate the neural correlates of learning with the wristband. We hypothesize that deaf participants who use the wristband will begin to show activation in their auditory cortex when feeling patterns of vibration on their wrist—in part because that will be the cortical territory available for takeover. Although we cannot have wearers participate in functional magnetic resonance imaging (fMRI) with the wristband (due to magnetic components in the device), we can either 1) investigate cortical activation with brain imaging techniques that do not rely on magnetic fields, such as functional near-infrared spectroscopy (fNIRS) (Ferrari and Quaresima, 2012) or 2) create an fMRI-safe version of the wristband that stimulates sensations on the wrist without needing magnetic components.

In conclusion, we have demonstrated the usefulness of an inexpensive, self-contained, wrist-worn device for the sensory substitution of sound.

6 Acknowledgements

We thank Scott Novich and several engineers from Neosensory for their background contributions to this work.

7 Conflicts of interest

All authors were employed by company Neosensory, Inc.

8 Author contributions

MVP and DME designed the experiments and wrote the manuscript. MVP and TA ran the experiments and analyzed the data.

9 References


10 Tables

Table 1. Demographic data. Hearing loss age indicates the age in years at which the participant started to lose hearing (or 0 if the participant was born deaf or with hearing loss). Lip reading self-rating indicates the level of proficiency in reading lips the participant reports, with 1 being “extremely bad” and 7 being “extremely good”. ASL usage indicates the amount the participant reports using American Sign Language (ASL) with friends, family, and colleagues, where 1 is “none at all” and 5 is “a great deal”. Hearing loss values are decibels of hearing loss at 6 pure tones in the left and the right ears. Note that 90 dB of hearing loss is the most amount of hearing loss the test can detect and indicates 90 dB or more of hearing loss.

<table>
<thead>
<tr>
<th></th>
<th>Age</th>
<th>Hearing loss age</th>
<th>Sex</th>
<th>Lip reading self-rating</th>
<th>ASL usage</th>
<th>Hearing loss (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>250 Hz</td>
</tr>
<tr>
<td>B1</td>
<td>23</td>
<td>4</td>
<td>F</td>
<td>4</td>
<td>5</td>
<td>90</td>
</tr>
<tr>
<td>B2</td>
<td>32</td>
<td>12</td>
<td>M</td>
<td>6</td>
<td>5</td>
<td>90</td>
</tr>
<tr>
<td>B3</td>
<td>35</td>
<td>3</td>
<td>F</td>
<td>5</td>
<td>1</td>
<td>90</td>
</tr>
<tr>
<td>B4</td>
<td>39</td>
<td>0</td>
<td>F</td>
<td>7</td>
<td>5</td>
<td>90</td>
</tr>
<tr>
<td>B5</td>
<td>42</td>
<td>0</td>
<td>M</td>
<td>5</td>
<td>5</td>
<td>90</td>
</tr>
<tr>
<td>B6</td>
<td>53</td>
<td>14</td>
<td>M</td>
<td>5</td>
<td>1</td>
<td>27</td>
</tr>
<tr>
<td>B7</td>
<td>54</td>
<td>45</td>
<td>F</td>
<td>5</td>
<td>2</td>
<td>14</td>
</tr>
<tr>
<td>B8</td>
<td>56</td>
<td>32</td>
<td>F</td>
<td>6</td>
<td>1</td>
<td>90</td>
</tr>
<tr>
<td>B9</td>
<td>58</td>
<td>6</td>
<td>F</td>
<td>7</td>
<td>2</td>
<td>90</td>
</tr>
<tr>
<td>B10</td>
<td>59</td>
<td>40</td>
<td>F</td>
<td>6</td>
<td>1</td>
<td>90</td>
</tr>
<tr>
<td>B11</td>
<td>59</td>
<td>5</td>
<td>F</td>
<td>1</td>
<td>3</td>
<td>90</td>
</tr>
<tr>
<td>B12</td>
<td>59</td>
<td>0</td>
<td>M</td>
<td>6</td>
<td>5</td>
<td>90</td>
</tr>
<tr>
<td>B13</td>
<td>62</td>
<td>4</td>
<td>M</td>
<td>7</td>
<td>5</td>
<td>90</td>
</tr>
<tr>
<td>B14</td>
<td>62</td>
<td>57</td>
<td>F</td>
<td>5</td>
<td>1</td>
<td>12</td>
</tr>
<tr>
<td>B15</td>
<td>66</td>
<td>3</td>
<td>M</td>
<td>4</td>
<td>4</td>
<td>90</td>
</tr>
<tr>
<td>B16</td>
<td>69</td>
<td>24</td>
<td>M</td>
<td>6</td>
<td>2</td>
<td>90</td>
</tr>
<tr>
<td>B17</td>
<td>80</td>
<td>17</td>
<td>F</td>
<td>7</td>
<td>3</td>
<td>90</td>
</tr>
<tr>
<td>B18</td>
<td>84</td>
<td>75</td>
<td>M</td>
<td>4</td>
<td>1</td>
<td>22</td>
</tr>
</tbody>
</table>

11 Figure Legends

Figure 1. The Neosensory Buzz. a. A wristband with 4 vibratory motors built into the strap. b. The sound captured by the Neosensory Buzz is played out through vibrating motors. A haptic illusion is
used such that the 4 motors are able to represent a continuous frequency space. The perceived
location of the vibration represents sound between 300 and 7500 Hz. The frequency space is scaled
logarithmically.

**Figure 2.** Sound Identification task. Sound files were algorithmically translated into patterns of
motor activation. When two motors are active, an illusory position in between their locations is
perceived. Five examples of 70 stimuli are shown.

**Figure 3.** Pattern Discrimination task. Spoken words are converted to sequences of motor activations
on the wristband. Three examples of 524 stimuli are shown.

**Figure 4.** Testing for Minimum Threshold - representative example. When presented with a train of
pulsing vibrations, participants pressed and held a button when the vibrations were strong enough to
feel and released the button when the vibrations were too weak to feel. The vibrations became weaker
after the button was pressed and stronger when it was released. The rate of change of the vibration
intensity decreased throughout the trial. The six red crosses indicate the values averaged to determine
a participant’s minimum threshold for a single trial.

**Figure 5.** Sound Identification performance.  
(A) Participants scored significantly higher on day 28
than they did on day 0 (p < .001). Box plots use Tukey’s original configuration (Mcgill, Tukey, and
Larsen 1978) with outliers defined as falling more than 1.5 times the interquartile range outside of
the box. (B) Participants who are older scored lower than their younger counterparts (p = .028).
Performance averaged over the 3 sessions. Shaded area shows a 95% confidence interval for the
linear regression.

**Figure 6.** Pattern Discrimination performance.  
(A) Participants scored significantly higher on non-
minimal pairs than they did on minimal pairs (p < .001). (B) Participants who are older tended to
score lower than their younger counterparts (non-significant, p = .075).

**Figure 7.** Minimum Threshold. Participants who are older tended to have a higher minimum
threshold intensity than their younger counterparts (non-significant, p = .122).
Figure 2

In review
Figure 3.

Minimal pair: "house"  
Non-minimal pair: "mouse"  
"when"

Translate to motor activations

Time

T10 ms
Figure 7: Performance (%) of Minimal vs. Non-minimal Pairs. The box plots show the distribution of performance with minimal pairs on the left and non-minimal pairs on the right. The dotted line represents the chance performance level. A significant difference is indicated by the *** symbol.