Enhanced Pitch Discrimination for Cochlear Implant Users with a New Haptic Neuroprosthetic

Mark D. Fletcher[1*], Nour Thini[2], Samuel W. Perry[1]

[1] University of Southampton Auditory Implant Service, University of Southampton, University Road, Southampton, SO17 1BJ, United Kingdom

[2] Faculty of Engineering and Physical Sciences, University of Southampton, University Road, Southampton, SO17 1BJ, United Kingdom

*corresponding author. Email: *M.D.Fletcher@soton.ac.uk*

Abstract

The cochlear implant (CI) is the most widely used neuroprosthesis, recovering hearing for more than half a million severely-to-profoundly hearing-impaired people. However, CIs still have significant limitations, with users having severely impaired pitch perception. Pitch is critical to speech understanding (particularly in noise), to separating different sounds in complex acoustic environments, and to music enjoyment. In recent decades, researchers have attempted to overcome shortcomings in CIs by improving implant technology and surgical techniques, but with limited success. In the current study, we take a new approach of providing missing pitch information through haptic stimulation on the forearm using our new mosaicOne_B device. The mosaicOne_B extracts pitch information in real-time and presents it via 12 motors that are arranged in ascending pitch along the forearm, with each representing a different pitch. In normal-hearing subjects listening to CI simulated audio, we showed that participants were able to discriminate pitch differences at a similar performance level to that achieved by normal-hearing listeners. Furthermore, the device was shown to be highly robust to background noise. This enhanced pitch discrimination has the potential to significantly improve music perception, speech recognition and speech prosody perception in CI users.

Key words: Touch perception, multi-sensory, cross-modal, somatosensory, music, hearing, auditory, electro-haptic stimulation

Introduction

The cochlear implant (CI) is a neuroprosthesis that allows hundreds of thousands of severely hearing-impaired people to hear again. To recover auditory perception, an array of micro-electrodes that deliver electrical pulses to the auditory nerve is surgically implanted into the cochlea. Due to anatomical and physical limitations, modern implants use only 12-24 electrodes to transfer sound information to the brain, although only around 8 electrodes are thought to be effective when used together[1,2]. In contrast, in a healthy cochlea, sound information is transferred to the auditory nerve by around 3500 hair cells[3]. Remarkably, despite these limitations, CIs allow the majority of users to identify words in quiet listening environments at an accuracy similar to those with normal hearing[4,5]. However, CI users are typically very poor at detecting pitch changes, which impairs their ability to identify age, sex, and accent[6,7], as well as perception of speech prosody[8-13]. Speech prosody allows a listener to distinguish statements from questions (e.g. “it is good.” from “it is good?”) and nouns from verbs (e.g. “**Ob**ject” from “ob**ject**”). It also allows listeners to distinguish emotion (e.g. anger from sadness) and intention (e.g. whether the phrase “nice jumper” was meant as a genuine complement or a sarcastic remark). Impaired pitch discrimination also limits music perception[14], as pitch conveys crucial melody, harmony, and tonality information. CI users struggle to recognise simple melodies[14-17] and to discriminate different instruments[14,18,19] with only around 13% of adult CI users reporting that they enjoy listening to music after implantation[20].

Traditionally, researchers and manufacturers have attempted to overcome the limitations of CIs by improving implant technology and surgical techniques. However, in recent decades, improvements in CI outcomes have slowed markedly[5,21]. In this study, we take a new approach. Rather than attempting to transfer more pitch information through the implant, we augment the electrical CI signal by delivering pitch information through haptic stimulation on the forearm (“electro-haptic stimulation”[22]). This approach is particularly appealing as this supplementary wearable neuroprosthetic is non-invasive and inexpensive.

The effectiveness of providing sensory information that is usually delivered through one sense using a different sense is well established. Seminal work by Paul Bach-y-rita in the late 1960s showed that, using visual information presented through tactile stimulation on the back, blind people can recognise faces, judge the speed and direction of an object, and complete complex inspection-assembly tasks[23,24]. Later, researchers successfully delivered visual information using sound[25,26] and basic speech information using haptic stimulation, either on the finger, forearm or wrist[27,28]. More recently, in addition to *substituting* auditory input for haptic input, it has been shown that it is possible to *augment* auditory input with haptic input; three recent studies have shown that CI users' ability to recognise speech in background noise was enhanced when speech information was presented through haptic stimulation on the wrists[22,29] or fingertips[30]. Two other recent studies have shown that haptic stimulation can improve melody identification in CI users, using a single-channel haptic stimulation device strapped to the fingertip[31] or wrist[32]. In the current study, we evaluated the ability of our new mosaicOne_B device, which delivers pitch information through multi-channel haptic stimulation along the forearm, to provide accurate pitch information.

The mosaicOne_B extracts pitch information from audio in real-time and delivers it through haptic stimulation. The device uses 12 motors, with six along the top and six along the underside of the forearm (see Figure 1). The motors are activated chromatically, like keys on a piano, with each motor representing a different pitch within a single octave. The mosaicOne_B delivers relative pitch information, meaning that sounds that are exactly an octave apart will produce the same pattern of stimulation. This approach allows for high relative pitch resolution, whilst discarding absolute pitch information (*i.e.* information on the pitch of a stimulus within the full scale of perceivable pitches). As even the poorest performing CI users are typically able to discriminate sounds that are an octave apart[33-36], by using the CI in combination with the mosaicOne_B*,* CI users are expected to have access to absolute pitch information.

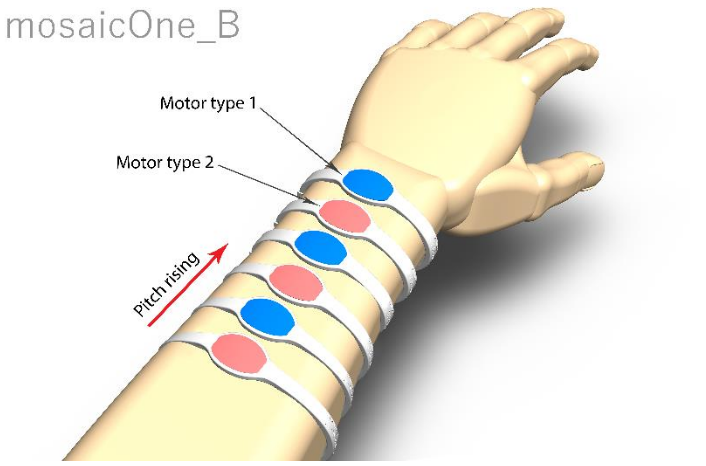

Figure 1: Schematic representation of the mosaicOne_B haptic stimulation device on the forearm. The two interleaved motor types used are represented by different colours.

The first aim of this study was to test the limits of pitch discrimination with the mosaicOne_B. Studies using real musical instruments have estimated average pitch discrimination thresholds across CI users of around 80-90% (i.e. 10-11 semitones)[35,36]. Other studies using synthetic sounds (tone complexes) have found average pitch discrimination thresholds of around 10-20% (2-3 semitones)[33,34]. In all studies, the variance across subjects was considerable, with some participants only able to discriminate sounds with a pitch difference of slightly less than 100% (one octave) and the best individuals able to discriminate around 3% (0.5 semitones). The performance of the best CI users is similar to pitch discrimination thresholds with musical instruments for normal-hearing listeners[35]. In the current study, we aimed to achieve an average pitch discrimination threshold of 6% (1 semitone) or better. This would allow CI users to track musical melodies (the smallest musical interval for western melodies is typically 1 semitone) and give access to cues for emotion and intention in speech.

Another way in which the mosaicOne_B could aid CI listening is by making it more robust to background noise. CI user performance degrades quickly when there are competing sounds; for example, CI users struggle to discriminate musical instruments when multiple instruments are playing[14,18,19] or to understand speech in noisy environments[21,22,30], such as classrooms, busy workplaces, or cafes. A number of studies have shown that, in addition to impairing speech prosody perception, reduced access to information about changes in speech fundamental frequency (F_0; an acoustic correlate of pitch) reduces speech recognition in noise[37,38]. The second aim of this study was to test whether the mosaicOne_B could provide accurate pitch information in the presence of background noise.

Finally, this study aimed to test whether pitch information from different modalities is combined effectively when delivered through audio and haptic stimulation, so that performance with audio and haptic stimulation together is better than with either alone. Alternatively, if one sense gives weak pitch information and the other strong pitch information, the weaker signal may create a distraction that impairs performance. There is a range of anatomical, physiological, and psychophysical evidence to suggest that audio and haptic signals are combined in the brain. Anatomical and physiological studies have revealed extensive connections between auditory and somatosensory neural pathways, from the periphery to cortex[39-42]. Psychophysical studies have demonstrated both that auditory stimuli can affect haptic perception[43] and that haptic stimuli can affect auditory perception[44-46]. In one study, it was shown that the perception of the dryness of a surface could be modulated by manipulating the accompanying audio[43]. In another set of studies, tactile stimulation was shown to increase perceived loudness and facilitate detection of faint sounds[45,46]. It may therefore be expected that pitch discrimination performance will be better when audio and haptic stimulation are provided concurrently than when either are presented alone.

In the twelve normal-hearing listeners tested in the current study, pitch discrimination was measured with CI-simulated audio alone, haptic stimulation alone, or with audio and haptic stimulation together. The stimuli were harmonic tone complexes that were designed to differ only in pitch (see Methods), so that the results can be generalised to both speech and musical sounds. For each of the three conditions, measurements were made with no background noise, and with background noise at signal-to-noise ratios (SNRs) of either –5 dB or –7.5 dB. These background noise levels were selected as assessment of pitch estimation errors produced by the mosaicOne_B increased at these SNRs (see Methods). It should be noted that these SNRs are far more challenging than those in which CI users are typically able to perform on speech-in-noise recognition tasks[21,22,30].

Results

Figure 2 shows pitch discrimination thresholds with audio stimulation only, with haptic stimulation only, and with audio and haptic stimulation together. Results are shown without background noise and with white background noise at either –5 dB or –7.5 dB SNR. A Friedman ANOVA was conducted with stimulation type (audio only, haptic only, audio-haptic) as a factor. A significant overall effect of stimulation type was found (χ^2 18.17, p = <.001). A significant overall effect of noise was also found for audio (χ^2 18.17, p = <.001) and haptic stimulation only (χ^2 15.45, p = .001). For audio only, pitch discrimination increased from a median change in F_0 of 43.4% without noise (ranging from 8.4% to 106.0% across participants) to 82.2% with noise at –5 dB SNR (ranging from 27.6% to 130%) and to 85.2% with noise at –7.5 dB SNR (ranging from 29.7% to 116.5%). For haptic only, median pitch discrimination was 1.4 % without noise (ranging from 0.8% to 3.5%), 2.0% with noise at –5 dB SNR (ranging from 0.6% to 6.6%), and 5.0% with noise at –7.5 dB SNR (ranging from 1.1% to 10.8%). No effect of noise was found for audio-haptic stimulation (χ^2 2.09, p = <.35). In the audio-haptic condition, median pitch discrimination thresholds were 1.5% without noise (ranging from 0.8% to 4.1%), 2.5% with noise at –5 dB SNR (ranging from 0.8% to 5.5%), and 2.4% with noise at –7.5 dB SNR (ranging from 0.9% to 15.0%).

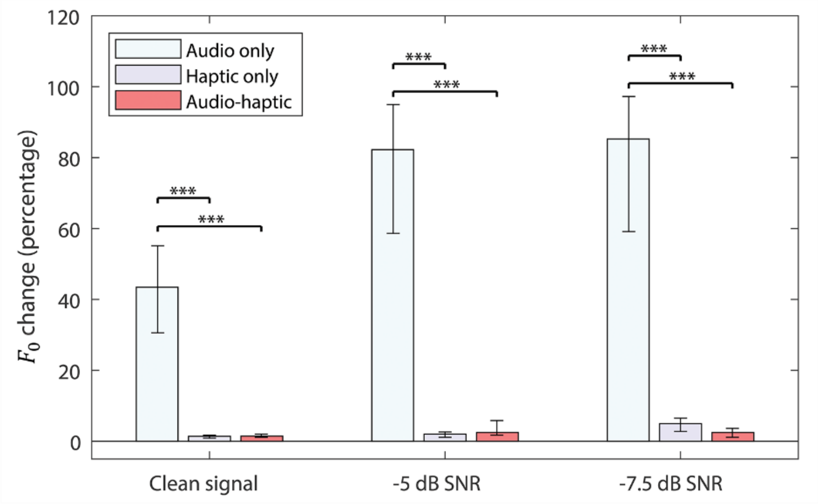

Figure 2: Median fundamental frequency (F_0) discrimination thresholds across our 12 participants with CI simulated audio only, haptic stimulation only, and audio and haptic stimulation together. Conditions with no background noise and with background noise at either –5 dB or –7.5 dB signal-to-noise ratio (SNR) are shown. Error bars show 5% and 95% confidence intervals (bootstrapped for each condition using 1000 samples with replacement).

Three post-hoc Wilcoxon signed-rank tests (with Holm-Bonferroni correction for multiple comparisons) were conducted to assess the effect of haptic stimulation on pitch discrimination thresholds (see Methods). Pitch discrimination was significantly better with audio-haptic stimulation than with audio alone (t = 78, p = .001, d = 3.76). Median discrimination thresholds improved by 42.0% without noise (ranging from 7.5% to 103.4% across participants), by 80.2% with noise at –5 dB SNR (ranging from 7.5% to 103.6%), and by 80.3% with noise at –7.5 dB SNR (ranging from 5.7% to 95.2%). Pitch discrimination was also significantly better with haptic alone than with audio alone (t = 78, p = .001, d = 3.75). Discrimination improved by 41.9% without noise (ranging from 7.2% to 104.9% across participants), by 79.8% with noise at –5 dB SNR (ranging from 7.0% to 101.0%), and by 80.8% with noise at –7.5 dB SNR (ranging from 7.1% to 91.0%). No difference in pitch discrimination was found between haptic alone and audio-haptic stimulation (t = 35, p = .791, d = -0.05).

Discussion

In this study, we found that the mosaicOne_B substantially improved pitch discrimination for normal-hearing subjects listening to CI simulated audio. The average pitch discrimination threshold with haptic stimulation was 1.4 % (without noise), which corresponds to markedly less than a quartertone, and is comfortably better than our target of 6% (1 semitone). Furthermore, even the worst performer in the current study was substantially better than our targeted average performance across subjects, achieving a pitch discrimination threshold of just 3.5% - comfortably less than a semitone (the minimum pitch change in most western melodies) and similar to pitch discrimination of the best performing CI users[33,34]. For both haptic alone and audio-haptic conditions, some participants achieved pitch discrimination thresholds as low as just 0.8%. This is similar to the performance of normal-hearing listeners for a similar auditory stimulus (although it should be noted that pitch discrimination thresholds for audio is highly sensitive to stimulus parameters)[47]. This enhanced pitch discrimination by the mosaicOne_B has the potential to significantly aid music perception in CI users, as well as speech recognition and speech prosody perception.

The excellent pitch discrimination performance found for the mosaicOne_B was more robust to background noise than may have been expected, with some participants achieving pitch discrimination thresholds of just 0.9% even when the noise was 7.5 dB louder than the signal. In fact, no effect of noise on pitch discrimination thresholds was found. At -7.5 dB SNR (the lowest used in the current study), even the best CI users are unable to perform pitch[48] or speech recognition tasks[21,22,30]. The absence of an effect of noise was surprising given the greater pitch estimation error by the mosaicOne_B at this low SNR, which led to a wider distribution of the motors being activated for a single stimulus (see Methods). It is possible that discrimination was achieved by a comparison of the time-averaged distributions of active motors for each stimulus. A similar process is thought to underlie signal detection in the auditory system[49,50]. The robustness to noise that was achieved by our real-time signal processing strategy is particularly impressive as pitch extraction algorithms tend to be highly susceptible to background noise and can often not be applied in real time[51]. It should be noted that, in the current study, the background noise used was non-harmonic (like environmental sounds such as rain or wind, but unlike competing talkers or background music). Future work is required to explore whether the current approach can also be successfully applied in environments with multiple harmonic sounds.

No difference in performance was found between the audio-haptic and haptic-alone conditions. The absence of a degradation in performance is encouraging, as it indicates that the poor-quality pitch information from auditory stimulation did not distract participants, even after only a small amount of familiarization. It is perhaps not surprising that performance with haptic stimulation was not enhanced by the addition of apparently much poorer pitch information provided through audio stimulation (as indicated by the better performance in the haptic-alone than audio-alone condition). Indeed, it has been observed in several previous studies that the greatest benefit from multisensory integration occurs when senses provide relatively low-quality information when used in isolation[52-54] (the principal of inverse effectiveness), which was not the case for haptic stimulation. Another reason for the absence of audio-haptic integration may have been the lack of training given. In previous studies, it has been shown that training is critical for audio-haptic integration. This has been shown for haptic enhancement of spatial hearing in CI users[55] and of speech recognition in noise both for CI users[22] and for normal-hearing listeners listening to CI simulated audio[29]. It is possible that audio-haptic integration would have been observed in the current study if training was provided.

It should be noted that the current study has not demonstrated that the auditory percept of pitch has been enhanced, but rather that participants were able to access higher-resolution pitch information through haptic stimulation. Participants in a previous study using haptic stimulation to enhance speech-in-noise performance[22] gave subjective reports that, after training, the speech sounded louder or clearer when haptic stimulation was provided. This indicates that haptic stimulation was able to modulate the audio percept. This idea is supported by psychophysical evidence (discussed in the introduction) that haptic stimulation can modulate auditory perception of loudness and the perception of aspirated and non-aspirated syllables[44-46]. Further work is required to establish whether the auditory percept of pitch can be modulated by haptic stimulation.

The average performance of our participants in the audio-only condition was consistent with previous studies of pitch discrimination in CI users[33-36]. However, previous studies did not use the precise stimulus used in the current study and average performance ranges markedly across studies. In the current study, there was a wide range of performance (of around one octave) across participants. This is also consistent with previous studies using CI users. It should be noted that, while performance of normal-hearing listeners listening to CI simulated audio matched that of CI users, the way that sounds were perceived may have differed.

There are limitations of the current study that should be noted. One such limitation is that pitch discrimination was only demonstrated for a reference signal with an F_0 close to 300 Hz (F_0 was roved). To ensure that the current approach could be applied to signals with different F_0s, mosaicOne_B outputs were assessed for several F_0s (see Methods). The outputs showed good consistency, which indicates that the results of the current study can be generalised to a range of F_0s. A second consideration is the age of the participants who took part (all of whom were under 32 years of age). A substantial portion of the CI user community is significantly older than the population tested in the current study. However, while spatial discrimination on the forearm is known to decline with age, the ability to distinguish between two stimulation points on the forearm remains less than the motor spacing for the mosaicOne_B (3 cm), even in older people[56]. Therefore, the findings of the current study are expected to be translatable to older populations. Finally, only a small amount of training was given in the current study, which may have led to pitch discrimination thresholds being underestimated. Previously, researchers have reported that, for normal-hearing listeners, auditory frequency discrimination performance continues to improve for around two weeks when two hours of training is given each day[57]. Furthermore, studies of enhancement of speech-in-noise performance with haptic stimulation for CI users have shown the importance of training for maximizing benefit[22,29]. Future work should assess whether training can lead to further enhancements in pitch discrimination with the mosaicOne_B*.*

Several steps are required to maximize the potential of the mosaicOne_B to bring real-world benefits to CI users. Firstly, it will be important to verify the findings of the current study in CI users. Future work should also seek to optimize the pitch extraction techniques used to reduce estimation errors in noise. Another important step, already discussed, is to assess the ability of the mosaicOne_B to effectively extract pitch cues in the presence of multiple harmonic sounds. Additionally, future studies should assess the effectiveness of the mosaicOne_B for improving speech perception, both in quiet and in noise, and music perception. Future developments to the mosaicOne_B could also include the exploitation of spatial hearing cues. CI users have poor access to spatial cues and are extremely poor at locating sounds[58], which can lead to impaired threat detection and sound source segregation. For example, it is well established that access to spatial hearing cues can enhance detection of signals, such as speech, in noise[59,60]. A recent study has shown strong evidence that haptic stimulation can be used to enhance localisation of sounds in CI users[55] and a similar approach might be implemented on the mosaicOne_B. Finally, a wearable neuroprosethic like the mosaicOne_B could aid CI users in everyday activities. It could incorporate additional features, such as a wake-up alarm (as CI users typically charge their implants during the night) and by connecting to smart devices in the Internet of Things, such as telephones, doorbells, ovens, and fire alarms.

The results of the current study demonstrate that the mosaicOne_B can extract and deliver precise pitch information through haptic stimulation. The device has been shown to be remarkably robust to non-harmonic background noise, which is common in real-world environments. The mosaicOne_B has several properties that make it suitable for a real-world application: stimulation was delivered to the forearm (a suitable site for a real-world use), the signal processing was performed in real-time, and the haptic signal was delivered using low-powered, compact motors. The findings of the current study suggest that the mosaicOne_B could offer a non-invasive and inexpensive means to improve speech and music perception in CI users.

Methods

Participants

Twelve participants (3 male and 9 female, aged between 22 and 31 years old) were recruited from the staff and students of the University of Southampton, and from acquaintances of the researchers. Participants gave written informed consent and no payment was given for participation. All participants reported no hearing or touch issues, had received no musical training, and did not speak a tonal language. Vibrotactile detection thresholds were measured at the fingertips of the left and right index fingers. Thresholds were measured at 31.5 Hz and 125 Hz, following conditions and criteria specified in ISO 13091-1:2001[61] (the fingertip was used as there are no published standards for normal wrist or forearm sensitivity). All participants had vibrotactile detection thresholds within the normal range (< 0.4 ms^{-2} RMS at 31.5 Hz, and < 0.7 ms^{-2} RMS at 125 Hz[61]), indicating no touch perception issues. Participants were also assessed by otoscopy and pure-tone audiometry. Participants had hearing thresholds not exceeding 20 dB hearing level (HL) at any of the standard audiometric frequencies between 0.25 and 8 kHz in either ear.

Stimuli

In testing and task familiarisation (see Procedure), the reference stimulus was a harmonic complex, with an average F_0 of 300 Hz (within the range of F_0s found for many musical instruments, approximately central to the range of F_0s found in human speech[62], and the frequency at which pitch cues are reduced for CI users[14]). The F_0 was roved by ±5% on each presentation (with a uniform distribution). The stimulus comprised of equal-amplitude harmonics generated up to 24 kHz (the Nyquist frequency). This signal was band-pass filtered between 1 kHz and 4 kHz, with 12th order (72 dB per octave) 0-phase Butterworth filters, to remove non-pitch cues (such as differences in the brightness of the sound[63], as discussed in Mehta and Oxenham[64]]). The signal had a duration of 500 ms, with 20 ms quarter-sine and -cosine onset and offset ramps. The target and reference stimuli were separated by 300 ms. The target and reference stimuli were the same, except that the F_0 of the target stimuli was adjusted following the adaptive track described in the procedure section. The level of the target and reference was nominally set to 65 dB SPL (RMS), but was roved on each presentation within a ±3 dB range (with a uniform distribution) to reduce potential loudness cues. The masking stimulus was a white noise, selected to equally mask each of the components of the harmonic complex.

The audio stimuli were processed using the SPIRAL vocoder to simulate CI listening. The SPIRAL vocoder is an advanced CI simulator that aims to bridge the gap between traditional tone- and noise-based simulations[65]. The SPIRAL was set to simulate 22 CI electrodes, with a current decay slope of 16 dB per octave, using 80 carrier tones. The test stimuli were delivered to the participants' right ear only. In the audio alone and audio-haptic conditions, pink noise at a level of 55 dB SPL was delivered to the left ear to mask any audio cues from the mosaicOne_B*.* In the haptic-alone condition, the pink noise was delivered to both ears.

In the mosaicOne_B familiarization app (see Procedure), two stimulus types were used. In both of the app's modules, CI simulation was not applied to the audio. In the pitch slider module, a constant tone was presented, and the frequency was adjusted between D3 and B3 on the chromatic scale based on the slider position. In the interval training module, two tones were presented. The tones were 500 ms long, with 20 ms quarter-sine and -cosine onset and offset ramps and were separated by a 100 ms gap. Frequencies were selected at random between D3 and B3 on the chromatic scale.

Tactile Signal Processing

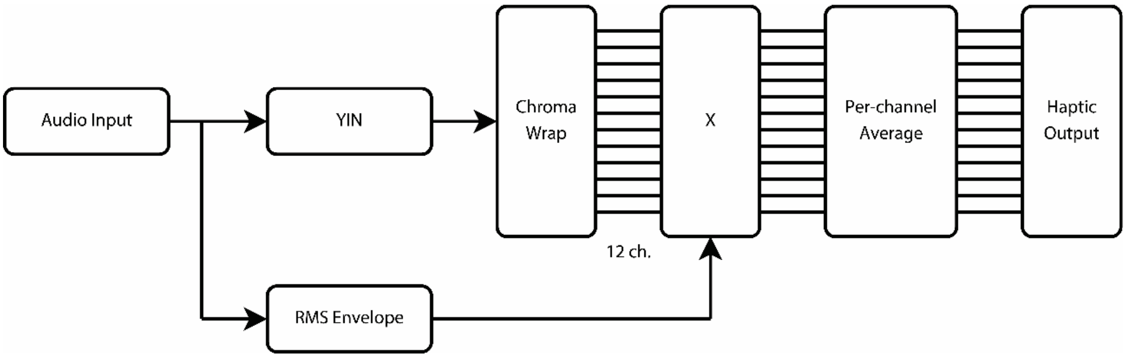

Figure 3: Schematic illustration of the signal processing chain for haptic signal generation.

Melody in music and prosody in speech is typically conveyed in sub-octave frequency shifts, with the absolute height of the pitch being largely irrelevant. In the current study, we used a pitch chroma analysis, which groups frequencies by octave to produce a spectral representation of relative pitch perception, discarding absolute pitch height information. A schematic representation of the signal processing chain that was used to convert audio to a tactile signal is illustrated in Figure 3. The haptic signal was generated by first estimating the F_0 and amplitude envelope of the input signal. F_0 was estimated using YIN, implemented in the Max Sound Box toolbox (version 2018-3, IRCAM, Paris, FR). A 14 ms window size was used (giving a minimum possible F_0 estimation of approximately 70 Hz) with no downsampling. The resulting F_0 estimate was then used to activate one of the 12 shakers on the mosaicOne_B. This was achieved by first mapping the F_0 to the MIDI scale, a commonly used scale for relating musical pitch to frequency. This representation was then split into 12 frequency channels, relative to a base frequency of 440 Hz. The full frequency mapping was defined as:

$$f_{\text{wrap}}[n] = \operatorname{mod}\left(69 + 12 \cdot \log_2\left(\frac{F_0[n]}{440}\right), 12\right),$$

$$y_i[n] = \begin{cases} 1, & i = f_{\text{wrap}}[n] \\ 0, & \text{otherwise} \end{cases},$$

where f_{wrap} is an integer in the range $0 \leq f_{\text{wrap}} < 12$, and y_i is the channel at index i.

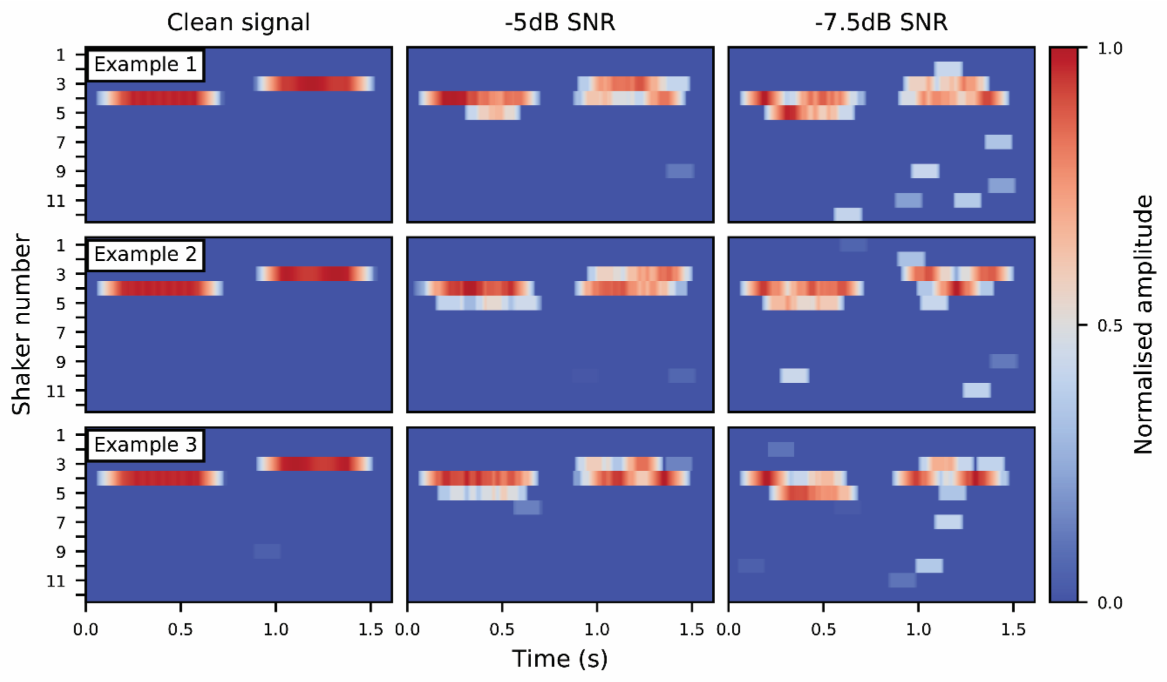

Figure 4: Haptic signals in response to three example harmonic complexes for the clean, -5 dB and -7.5 dB SNR conditions. The reference signal is followed by a target signal with an F_0 increased by 5%.

The RMS amplitude envelope of the input audio was calculated in parallel, using a 14 ms window. The activated channel was then multiplied by this envelope. Finally, a moving RMS average of each of the 12 channels was calculated using a 125 ms window. This per-channel averaging acted as a simple noise-reduction method. This helped reduce the effects of artefacts produced by the F_0 estimation algorithm as background noise increased. The haptic output in response to the harmonic complexes (described in the Stimuli section) is illustrated in Figure 4.

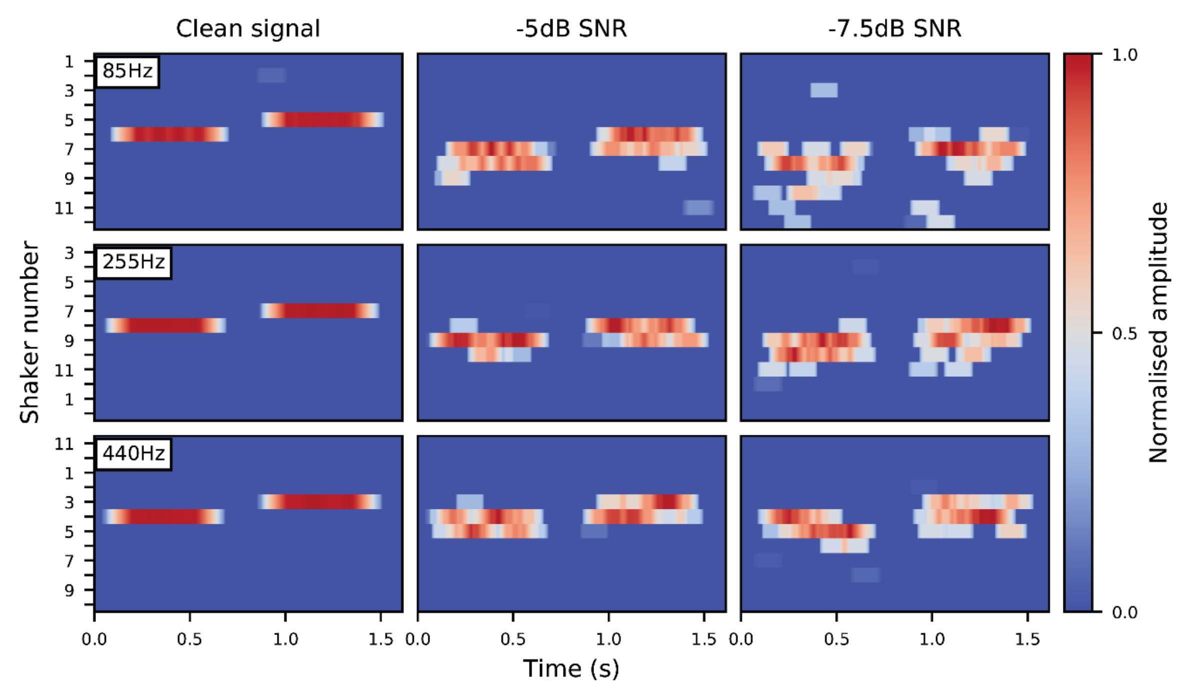

Figure 5: Haptic signals in response to a harmonic complex (sawtooth wave) at 85 Hz, 255 Hz and 440 Hz in clean, -5dB and -7.5dB SNR conditions. The reference signal is followed by a target signal with an F_0 increased by 5%. Shaker numbers are offset to centre the reference stimulus for display purposes.

The performance of the algorithm was assessed for different stimulus frequencies. Sawtooth waves at 85 Hz (lowest average F_0 of typical male speech[62]), 255 Hz (highest average F_0 of typical female speech[62]) and 440 Hz (standard tuning pitch for western music). A sawtooth harmonic complex was used as it consists of equivalent odd and even harmonics (that decrease in amplitude with increasing frequency – as is typical of real-world stimuli, such as speech). Bandpass filtering was not applied as removal of non-pitch cues was not necessary. Figure 5 illustrates the algorithm's performance. Performance at decreasing SNRs is comparable to the test stimulus, with marginally poorer performance at 85 Hz for the -7.5 dB SNR condition. Additionally, for the -5 and -7.5 dB SNR conditions at 85 Hz and 255 Hz, estimates are offset by 1-2 shakers relative to the clean condition. These errors are due to the inaccuracy of initial F_0 estimation, and the non-linear mapping of frequency (which requires greater precision of F_0 estimation at lower frequencies). Despite these errors, relative pitch differences appear largely unaffected.

Apparatus

During pure-tone audiometry, participants were seated in a sound-attenuated booth with a background noise level conforming to British Society of Audiology recommendations[66].

Audiometric measurements were conducted using a Grason-Stadler GSI 61 Clinical Audiometer and Telephonics 296 D200-2 headphones. Vibro-tactile threshold measurements were made using a HVLab Vibro-tactile Perception Meter with a 6-mm contactor that had a rigid surround and a constant upward force of 2N (following International Organization for Standardization specifications[61]). This system was calibrated using a Bruel & Kjaer (B&K) calibration exciter (Type 4294).

The experiment took place in a quiet listening room. During testing, the experimenter sat behind a screen with no line of site to the participant. The participants responded by pressing buttons on a iiyama ProLite T2454MSC-B1AG 24-inch touchscreen monitor. All stimuli were generated using custom MATLAB scripts (version R2019a, The MathWorks Inc., Natick, MA, USA) and controlled using Max 8 (version 8.0.8). Both audio and haptic signals were played out at a sample rate of 48 kHz via a MOTU 24Ao soundcard (MOTU, Cambridge, MA, USA). Audio was presented using ER-2 insert earphones (Etymotic, IL, USA) and the haptic signal was delivered through the mosaicOne_B via the mosaicOne_B haptic interface (for amplification of haptic signals). Audio stimuli were calibrated using a B&K G4 sound level meter, with a B&K 4157 occluded ear coupler (Royston, Hertfordshire, UK). Sound level meter calibration checks were carried out using a B&K Type 4231 sound calibrator.

Haptic stimulation was delivered using the mosaicOne_B (Figure 1 shows a schematic representation of the device). The mosaicOne_B had twelve motors, with six strapped to the top and six to the bottom of the forearm. The motors were attached using six elastic straps, fastened with Velcro. Two motor types, the Precision Microdrives 304-116 5 mm vibration motor (labelled “Motor type 1” in Figure 1) and the Precision Microdrives 306-10H 7 mm vibration motor (labelled “Motor type 2” in Figure 1) were used in an interleaved fashion, with each motor separated by 3 cm. The bottom motors were arranged in reverse order to maximize the distance between motor types. The motors were calibrated so that the driving signal extrema corresponded to the output amplitude extrema (maximum amplitudes of 1 and 1.84 G, respectively). This maximised the dynamic range of the motors. The different motor types have different operating frequencies of 280 Hz for the 5 mm motor and 230 Hz for the 7 mm motor. This configuration was selected to maximize differentiation between motors by allowing the user to exploit both location and frequency cues. The different motors were expected to be discriminable in frequency based on frequency discrimination of the skin[67]. The motors were also expected to be spatially discriminable, even in older users, based on two-point discrimination thresholds[56]. Note that it has been argued that two-point discrimination thresholds likely over-estimate the minimum location separation required to discriminate motors[24]. This suggestion was supported by informal testing during development of the mosaicOne_B.

Procedure

The experiment had three phases, all of which were completed in a single session lasting around two hours. The first phase was the screening phase. During screening, participants first completed a questionnaire to ensure that they (1) had no conditions or injuries that may affect their touch perception, (2) had not been exposed to sustained periods of intense hand or arm vibration at any time, (3) had no recent exposure to hand or arm vibration, (4) had no conditions or injuries that may affect their hearing perception, (5) had received no musical training at any time, or (6) did not speak a tonal language. Next, audiometric hearing thresholds were measured to ensure participants had normal hearing (thresholds <20 dB HL). Thresholds were measured following British Society of Audiology guidelines[66]. Following this, vibro-tactile detection thresholds were measured at the fingertip, to check for normal touch perception (< 0.4 ms^{-2} RMS at 31.5 Hz, and < 0.7 ms^{-2} RMS at 125 Hz[61]). Thresholds were measured following the protocol recommended by the International Organization for Standardization[61]. Finally, otoscopy was performed to ensure insert earphones could safely be used. If the participant passed all screening stages, they continued to the familiarization phase.

In the familiarization phase, participants first used an app developed to familiarize them with the mosaicOne_B. Participants used the app for 5-10 minutes and were invited to ask questions if anything was unclear. The app consisted of a pitch slider and an interval training module. For each module, participants could switch between haptic only, audio-haptic and audio only modes. In both modules, CI simulation was not applied to the audio. In the pitch slider module, a constant tone was played, and the frequency of the tone was adjusted based on slider position. In the interval training module, participants could select either a “Low → High” or “High → Low” button, which determined the pitches of two consecutive tones. The number of presentations was not limited, but any given presentation could not be repeated.

After using the app to familiarize themselves with the device, participants were familiarized with the task used in the testing phase. Participants completed a short practice session of 15 trials for each condition. In the testing and task familiarization, a two-alternative forced-choice task was used in which participants were asked to judge which interval contained the sound or vibration stimulus with the higher pitch. Participants used two buttons labelled “1” and “2” to select whether the first or second stimulus was higher in pitch. Visual feedback was given, indicating whether the response was correct or incorrect. The pitch difference between intervals was initially set at +80% of the reference pitch and was then varied using a one-up, two-down adaptive procedure, with percentage difference varying by 10% for the first two reversals, 5% for the third reversal and 1% for the remaining four reversals. Thresholds for each track were calculated as the mean of the last four reversals. The order of conditions (audio only, audio-haptic, and haptic only) was counterbalanced across participants, and the noise conditions (no noise, noise at –5 dB SNR, and noise at –7.5 dB SNR) were presented in a random order for each condition.

The experimental protocol was approved by the University of Southampton Faculty of Engineering and Physical Sciences Ethics Committee (ERGO ID: 47769). All research was performed in accordance with the relevant guidelines and regulations.

Statistics

The data were analysed using a Friedman ANOVA (with Bonferoni-Holm correction for multiple comparisons). Three planned post-hoc Wilcoxon signed-rank tests were also performed (also Bonferoni-Holm corrected). Non-parametric tests were used as the data was not normally distributed.

Data Availability

The dataset from the current study is publicly available through the University of Southampton’s Research Data Management Repository (DOI will be provided by the university if this manuscript is accepted).

Acknowledgements

Thank you to Carl Verschuur for his consistent support and encouragement. Thank you also to Anahita Mehta for extremely helpful input to the design of this study, to Eric Hamdan and P.T. Chaplin for technical support, to Jana Zgheib (Knárc Bön) for insightful comments during the piloting of this study, and to Ben Lineton for statistics support. Our deepest gratitude also goes to Alex Fletcher and Helen Fletcher, whose patience and calmness facilitated the writing of this manuscript. Finally, we are grateful to each of the participants who gave up their time to take part in this study.

Salary support for author MDF was provided by The Oticon Foundation. Salary support for author SWP was provided by The Oticon Foundation and the Engineering and Physical Sciences Research Council (UK).

Author Contributions

MDF and SWP designed and implemented the experiment. NT, MDF, and SWP conducted the experiment. MDF and SWP wrote the manuscript text. All authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

References

1 Wilson, B. S. & Dorman, M. F. Cochlear implants: a remarkable past and a brilliant future. *Hear Res* **242**, 3-21, (2008).

2 Friesen, L. M., Shannon, R. V., Baskent, D. & Wang, X. Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants. *J Acous Soc Am* **110**, 1150-1163, (2001).

3 Moller, A. R. Hearing: it's physiology and pathophysiology. *New York: Academic Press* (2000).

4 Fetterman, B. L. & Domico, E. H. Speech recognition in background noise of cochlear implant patients. *Otolaryngol Head Neck Surg* **126**, 257-263, (2002).

5 Zeng, F. G., Rebscher, S., Harrison, W., Sun, X. & Feng, H. Cochlear implants: system design, integration, and evaluation. *IEEE Rev Biomed Eng* **1**, 115-142, (2008).

6 Abberton, E. & Fourcin, A. J. Intonation and speaker identification. *Lang Speech* **21**, 305-318, (1978).

7 Titze, I. R. Physiologic and Acoustic Differences between Male and Female Voices. *J Acoust Soc Am* **85**, 1699-1707, (1989).

8 Banse, R. & Scherer, K. R. Acoustic profiles in vocal emotion expression. *J Pers Soc Psychol* **70**, 614-636, (1996).

9 Murray, I. R. & Arnott, J. L. Toward the Simulation of Emotion in Synthetic Speech - a Review of the Literature on Human Vocal Emotion. *J Acoust Soc Am* **93**, 1097-1108, (1993).

10 Most, T. & Peled, M. Perception of suprasegmental features of speech by children with cochlear implants and children with hearing aids. *J Deaf Stud Deaf Edu* **12**, 350-361, (2007).

11 Peng, S. C., Tomblin, J. B. & Turner, C. W. Production and perception of speech intonation in pediatric cochlear implant recipients and individuals with normal hearing. *Ear Hearing* **29**, 336-351, (2008).

12 Meister, H., Landwehr, M., Pyschny, V., Walger, M. & von Wedel, H. The perception of prosody and speaker gender in normal-hearing listeners and cochlear implant recipients. *Int J Audiol* **48**, 38-48, (2009).

13 Xin, L., Fu, Q. J. & Galvin, J. J., 3rd. Vocal emotion recognition by normal-hearing listeners and cochlear implant users. *Trends Amplif* **11**, 301-315, (2007).

14 McDermott, H. J. Music perception with cochlear implants: a review. *Trends Amplif* **8**, 49-82, (2004).

15 Kong, Y. Y., Cruz, R., Jones, J. A. & Zeng, F. G. Music perception with temporal cues in acoustic and electric hearing. *Ear Hear* **25**, 173-185, (2004).

16 Galvin, J. J., 3rd, Fu, Q. J. & Nogaki, G. Melodic contour identification by cochlear implant listeners. *Ear Hear* **28**, 302-319, (2007).

17 Zeng, F. G., Tang, Q. & Lu, T. Abnormal pitch perception produced by cochlear implant stimulation. *PLoS One* **9**, (2014).

18 Gfeller, K., Witt, S., Woodworth, G., Mehr, M. A. & Knutson, J. Effects of frequency, instrumental family, and cochlear implant type on timbre recognition and appraisal. *Ann Otol Rhinol Laryngol* **111**, 349-356, (2002).

19 Prentiss, S. M. *et al.* Temporal and spectral contributions to musical instrument identification and discrimination among cochlear implant users. *World J Otorhinolaryngol Head Neck Surg* **2**, 148-156, (2016).

20 Philips, B. *et al.* Characteristics and determinants of music appreciation in adult CI users. *Eur Arch Otorhinolaryngol* **269**, 813-821, (2012).

21 Wilson, B. S. Getting a decent (but sparse) signal to the brain for users of cochlear implants. *Hear Res* **322**, 24-38, (2015).

22 Fletcher, M. D., Hadeedi, A., Goehring, T. & Mills, S. R. Electro-haptic enhancement of speech-in-noise performance in cochlear implant users. *Sci Rep* **9**, (2019).

23 Bach-y-Rita, P., Collins, C. C., Saunders, F. A., White, B. & Scadden, L. Vision substitution by tactile image projection. *Nature* **221**, 963-964, (1969).

24 Bach-y-Rita, P. Tactile sensory substitution studies. *Ann N Y Acad Sci* **1013**, 83-91, (2004).

25 Capelle, C., Trullemans, C., Arno, P. & Veraart, C. A real-time experimental prototype for enhancement of vision rehabilitation using auditory substitution. *IEEE Trans Biomed Eng* **45**, 1279-1293, (1998).

26 Meijer, P. B. L. An Experimental System for Auditory Image Representations. *Ieee T Bio-Med Eng* **39**, 112-121, (1992).

27 Hnath-Chisolm, T. & Kishon-Rabin, L. Tactile Presentation of Voice Fundamental-Frequency as an Aid to the Perception of Speech Pattern Contrasts. *Ear Hearing* **9**, 329-334, (1988).

28 Thornton, A. R. D. & Phillips, A. J. A comparative trial of four vibrotactile aids. *in Tactile Aids for the Hearing Impaired, edited by I. R. Summers*, Whurr, London, pp. 231–251. (1992).

29 Fletcher, M. D., Mills, S. R. & Goehring, T. Vibro-Tactile Enhancement of Speech Intelligibility in Multi-talker Noise for Simulated Cochlear Implant Listening. *Trends Hear* **22**, (2018).

30 Huang, J., Sheffield, B., Lin, P. & Zeng, F. G. Electro-Tactile Stimulation Enhances Cochlear Implant Speech Recognition in Noise. *Sci Rep* **7**, (2017).

31 Huang, J., Lu, T., Sheffield, B. & Zeng, F.-G. Electro-Tactile Stimulation Enhances Cochlear-Implant Melody Recognition. *Ear Hearing*, (2019).

32 Luo, X. & Hayes, L. Vibrotactile Stimulation Based on the Fundamental Frequency Can Improve Melodic Contour Identification of Normal-Hearing Listeners With a 4-Channel Cochlear Implant Simulation. *Front Neurosci* **13**, 1145, (2019).

33 Drennan, W. R. *et al.* Clinical evaluation of music perception, appraisal and experience in cochlear implant users. *Int J Audiol* **54**, 114-123, (2015).

34 Kang, R. *et al.* Development and validation of the University of Washington Clinical Assessment of Music Perception test. *Ear Hear* **30**, 411-418, (2009).

35 Brockmeier, S. J. *et al.* The MuSIC perception test: a novel battery for testing music perception of cochlear implant users. *Cochlear Implants Int* **12**, 10-20, (2011).

36 Bruns, L., Murbe, D. & Hahne, A. Understanding music with cochlear implants. *Sci Rep* **6**, (2016).

37 Laures, J. S. & Weismer, G. The effects of a flattened fundamental frequency on intelligibility at the sentence level. *J Speech Lang Hear Res* **42**, 1148-1156, (1999).

38 Binns, C. & Culling, J. F. The role of fundamental frequency contours in the perception of speech against interfering speech. *J Acoust Soc Am* **122**, 1765-1776, (2007).

39 Meredith, M. A. & Allman, B. L. Single-unit analysis of somatosensory processing in the core auditory cortex of hearing ferrets. *Eur J Neurosci* **41**, 686-698, (2015).

40 Shore, S. E., Vass, Z., Wys, N. L. & Altschuler, R. A. Trigeminal ganglion innervates the auditory brainstem. *J Comp Neurol* **419**, 271-285, (2000).

41 Aitkin, L. M., Kenyon, C. E. & Philpott, P. The representation of the auditory and somatosensory systems in the external nucleus of the cat inferior colliculus. *J Comp Neurol* **196**, 25-40, (1981).

42 Shore, S. E., El Kashlan, H. & Lu, J. Effects of trigeminal ganglion stimulation on unit activity of ventral cochlear nucleus neurons. *Neuroscience* **119**, 1085-1101, (2003).

43 Jousmaki, V. & Hari, R. Parchment-skin illusion: sound-biased touch. *Curr Biol* **8**, (1998).

44 Gick, B. & Derrick, D. Aero-tactile integration in speech perception. *Nature* **462**, 502-504, (2009).

45 Gillmeister, H. & Eimer, M. Tactile enhancement of auditory detection and perceived loudness. *Brain Res* **1160**, 58-68, (2007).

46 Schurmann, M., Caetano, G., Jousmaki, V. & Hari, R. Hands help hearing: Facilitatory audiotactile interaction at low sound-intensity levels. *J Acoust Soc Am* **115**, 830-832, (2004).

47 Kaernbach, C. & Bering, C. Exploring the temporal mechanism involved in the pitch of unresolved harmonics. *J Acoust Soc Am* **110**, 1039-1048, (2001).

48 Kreft, H. A., Nelson, D. A. & Oxenham, A. J. Modulation Frequency Discrimination with Modulated and Unmodulated Interference in Normal Hearing and in Cochlear-Implant Users. *J Assoc Res Oto* **14**, 591-601, (2013).

49 Verhey, J. L., Rennies, J. & Ernst, S. M. A. Influence of envelope distributions on signal detection. *Acta Acust United Ac* **93**, 115-121 (2007).

50 McDermott, J. H., Schemitsch, M. & Simoncelli, E. P. Summary statistics in auditory perception. *Nat Neurosci* **16**, 493-498, (2013).

51 Jouvet, D. & Laprie, Y. Performance Analysis of Several Pitch Detection Algorithms on Simulated and Real Noisy Speech Data. *Eur Signal Pr Conf*, 1614-1618 (2017).

52 Laurienti, P. J., Burdette, J. H., Maldjian, J. A. & Wallace, M. T. Enhanced multisensory integration in older adults. *Neurobiol Aging* **27**, 1155-1163, (2006).

53 Hairston, W. D., Laurienti, P. J., Mishra, G., Burdette, J. H. & Wallace, M. T. Multisensory enhancement of localization under conditions of induced myopia. *Exp Brain Res* **152**, 404-408, (2003).

54 Wallace, M. T., Wilkinson, L. K. & Stein, B. E. Representation and integration of multiple sensory inputs in primate superior colliculus. *Journal of Neurophysiology* **76**, 1246-1266 (1996).

55 Fletcher, M. D., Cunningham, R. O. & Mills, S. R. Electro-Haptic Enhancement of Spatial Hearing in Cochlear Implant Users. *Sci Rep* **10**, 1621, (2020).

56 Leveque, J. L., Dresler, J., Ribot-Ciscar, E., Roll, J. P. & Poelman, C. Changes in tactile spatial discrimination and cutaneous coding properties by skin hydration in the elderly. *J Invest Dermatol* **115**, 454-458, (2000).

57 Moore, B. C. Frequency difference limens for short-duration tones. *J Acoust Soc Am* **54**, 610-619, (1973).

58 Dorman, M. F., Loiselle, L. H., Cook, S. J., Yost, W. A. & Gifford, R. H. Sound Source Localization by Normal-Hearing Listeners, Hearing-Impaired Listeners and Cochlear Implant Listeners. *Audiol Neurootol* **21**, 127-131, (2016).

59 Bronkhorst, A. W. & Plomp, R. The effect of head-induced interaural time and level differences on speech intelligibility in noise. *J Acoust Soc Am* **83**, 1508-1516, (1988).

60 Dirks, D. D. & Wilson, R. H. The effect of spatially separated sound sources on speech intelligibility. *J Speech Hear Res* **12**, 5-38, (1969).

61 International Organization for Standardization. Mechanical vibration–Vibrotactile perception thresholds for the assessment of nerve dysfunction–Part 1: Methods of measurement at the fingertips. *ISO 13091-1:2001* (2001).

62 Hess, W. Pitch Determination of Speech Signals: Algorithms and Devices. *Springer-Verlag* (1983).

63 Shackleton, T. M. & Carlyon, R. P. The role of resolved and unresolved harmonics in pitch perception and frequency modulation discrimination. *J Acoust Soc Am* **95**, 3529-3540, (1994).

64 Mehta, A. H. & Oxenham, A. J. Vocoder Simulations Explain Complex Pitch Perception Limitations Experienced by Cochlear Implant Users. *J Assoc Res Otolaryngol* **18**, 789-802, (2017).

65 Grange, J. A., Culling, J. F., Harris, N. S. L. & Bergfeld, S. Cochlear implant simulator with independent representation of the full spiral ganglion. *J Acoust Soc Am* **142**, (2017).

66 British Society of Audiology. Recommended procedure: Pure-tone air-conduction and bone-conduction threshold audiometry with and without masking. *British Society of Audiology*, 1-36, (2017).

67 Rothenberg, M., Verrillo, R. T., Zahorian, S. A., Brachman, M. L. & Bolanowski, S. J., Jr. Vibrotactile frequency for encoding a speech parameter. *J Acoust Soc Am* **62**, 1003-1012, (1977).