Efficient Evaluation of Coding Strategies for Transcutaneous Language Communication

Robert Turcott, Jennifer Chen, Pablo Castillo, Brian Knott, Wahyudinata Setiawan, Forrest Briggs, Keith Klumb, Freddy Abnousi, Prasad Chakka, Frances Lau, Ali Israr

Facebook, Inc, Menlo Park, CA 94025, USA
{rturcott, jenniferchen, pablocastillo, brianknott, wahyudinata, fbriggs, kklumb, abnousi, prasad, flau, aliisrar}@fb.com

Abstract. Communication of natural language via the skin has seen renewed interest with the advent of mobile devices and wearable technology. Efficient evaluation of candidate haptic encoding algorithms remains a significant challenge. We present 4 algorithms along with our methods for evaluation, which are based on discriminability, learnability, and generalizability. Advantageously, mastery of an extensive vocabulary is not required. Haptic displays used 16 or 32 vibrotactile actuators arranged linearly or as a grid on the arm. In Study 1, a two-alternative, forced-choice protocol tested the ability of 10 participants to detect differences in word pairs encoded by 3 acoustic algorithms: Frequency Decomposition (FD), Dominant Spectral Peaks (DSP), and Autoencoder (AE). Detection specificity was not different among the algorithms, but sensitivity was significantly worse with AE than with FD or DSP. Study 2 compared the performance of 16 participants randomized to DSP vs a phoneme-based algorithm (PH) using a custom video game for training and testing. The PH group performed significantly better at all test stages, and showed better recognition and retention of words along with evidence of generalizability to new words.

Keywords: haptic communication, speech to touch, vibrotactile display, learning, phoneme, frequency decomposition, vocoder

1 Introduction

Transcutaneous language communication (TLC) allows the user to perceive and understand a tactile representation of spoken or written language [1–4]. In contrast to abstracted semantic content (e.g., emojis), TLC provides an unabbreviated, 1:1 translation between language and haptic stimulation, thereby preserving the richness and complexity of natural language.

TLC was an active area of research through the 1980s with a focus on aiding communication for the hearing impaired [5]. Momentum in the field waned with the development of the cochlear implant [6], but TLC has seen a renewed interest with the proliferation of mobile devices and wearable technology. By using the skin, TLC avoids competing with traditional visual and acoustic communication tools. In addition, it has the potential to provide a discreet, unobtrusive channel for receiving information.

Developing a successful TLC system faces several challenges, including matching mechanical stimulation to the psychophysical properties of the sensory system, designing a display that is sufficiently miniaturized and comfortable to be practical for extended use, establishing a learning process that is rapid and effective, and developing methods to translate language into stimulus patterns that allow real-time interpretation without an undue cognitive burden.

Even when translation algorithms are informed by psychophysics, the task of empirically evaluating candidate algorithms remains, and is particularly challenging because the intended applications typically require extensive training [6, 7]. The development of efficient methods of algorithm evaluation is a focus of this paper. Discriminability, learnability, and generalizability are identified as key criteria that can be evaluated without requiring mastery of a complex and extensive haptic vocabulary. These criteria were applied to 4 candidate algorithms, and provide an empirical basis for selecting the Phonemic algorithm for subsequent development.

The paper is organized as follow: Section 2 (Methods) presents a description of the haptic displays, translation algorithms, evaluation criteria, and protocols for Study 1 (discriminability) and Study 2 (learning). Section 3 presents the results of the studies and Section 4 discusses them. Finally, the conclusions are summarized in Section 5.

2 Methods

2.1 Haptic Display and Control

Each haptic display used 8 voice coils (VCs) (Tectonic Elements TEAX13C02-8/RH) in either a linear arrangement (25 mm between centers) or a 2x4 grid pattern (approximately 50 mm between centers). The VCs were secured to a 23-cm long thermoformed arced surface that served as a rigid backer. They were tailored for arm placement because of the convenience, relatively large surface area, and reasonable discrimination and sensitivity to vibration of this area (Fig. 1).

Fig. 1. Voice coil actuators were arranged linearly for the acoustic algorithms (left-hand panel) or in a grid pattern for the phonemic algorithm (middle panel). Displays were secured to one or both arms, depending on the protocol, using Velcro straps (right-hand panel).

The VCs were driven using audio .wav files via daisy-chained Motu 24Ao USB audio interfaces (Mark of the Unicorn, Cambridge, USA) with custom amplifier boards using the MAX98306 Stereo Amplifier (Adafruit Industries, New York).

2.2 Algorithms

We studied algorithms that convert language to haptic stimuli from 2 broad classes. Acoustic algorithms process recordings of spoken language to yield haptic control signals [2, 4]. The input is fundamentally acoustic, so changes in speech rate or pitch, for example, result in changes in the haptic stimuli. In contrast, phonemic algorithms first convert language to a sequence of phonemes, then map the phonemes to pre-defined haptic patterns. Features such as pitch, rate, and prosody are not intrinsic to the encoding, but can be conveyed as metadata and used as modifiers to the patterns. The algorithms we investigated are described below. The first 3 are acoustic, and the 4th is phonemic.

Frequency decomposition (FD). The speech waveform was digitally sampled at 44.1 kHz for Study 1 and 22.05 kHz for Study 2. It was passed through a pre-emphasis filter with α=0.97 before being partitioned into frames of 1024 samples (23.2 ms) with 50% overlap [8]. After multiplication by a Hamming window, the magnitude spectrum was obtained using the fast Fourier transform (FFT). 32 triangular-shaped filters with 50% overlap and unity area were created as follows: the spectrum between 100 Hz and 1 kHz was divided into 3 linearly spaced bands, and between 1 kHz and 8 kHz was divided into 29 mel-frequency-spaced bands [8]. The scaler product of the spectrum and each filter modulated a 200 Hz sine wave at the corresponding haptic actuator. The output was normalized such that the peak amplitude over the duration of the recording corresponded to 25-30 dB sensation level. Frequencies were mapped progressively from left wrist (low) to left shoulder, and from right wrist to right shoulder (high).

Dominant Spectral Peaks (DSP). This algorithm was identical to FD, except that for each frame, only the 5 largest filter outputs were used; all others were set to zero. For phonemes with a prominent formant structure (vowels) this yielded stimulation by actuators that corresponded to the frequency location of the dominant formants. For phonemes lacking formants (e.g., fricatives) this typically resulted in stimulation that migrated in location over the duration of the phoneme.

Autoencoder (AE). An autoencoder is an artificial neural network architecture that uses unsupervised learning to develop an efficient (compressed) representation of the input data [9]. Our embodiment used 128-unit input and output layers, and a bottleneck layer of 32 units, implemented in collaboration with Scyfer B.V. (Amsterdam, NL). Complex 128-point FFTs of recorded speech (down sampled from 22.05 to 16 kHz) with 25% overlap served as input data.

To improve the recognizability of the haptic patterns by humans, the following constraints were incorporated into the training cost function: sparsity (minimize the number

of active actuators), discreteness (quantize outputs to 1 of 3 states), and ordinality (sequence output levels rather than treat as categorical). The encodings represented in compressed form by the 32 trinary units of the hidden layer were used to drive the actuators, which were off, or driven with 100 Hz or 250 Hz sinusoidal activation at 25-30 dB sensation level, depending on the state of the unit.

Phonemic (PH). Motivated by findings of previous work [10], we implemented custom haptic patterns that were as simple as possible while still allowing discriminability. Sequences of phonemes were manually transcribed. Encoding rules included assigning dynamic, longer duration patterns to vowels, and approximating place of articulation for consonants with location on the arm (Fig. 2). Active actuators were implemented with voice coils driven at 250 Hz with amplitude 25-30 dB sensation level. Static patterns (consonants) and dynamic patterns (vowels) had durations of 120 and 220 ms, respectively. For dynamic patterns, individual actuators were on for 97 ms, with successive activations offset by 62 ms to enhance the illusion of continuous motion [11]. Activation onset and offset were ramped over a duration that was 10% of the on-time. A 200 ms inter-stimulus interval was used between phonemes.

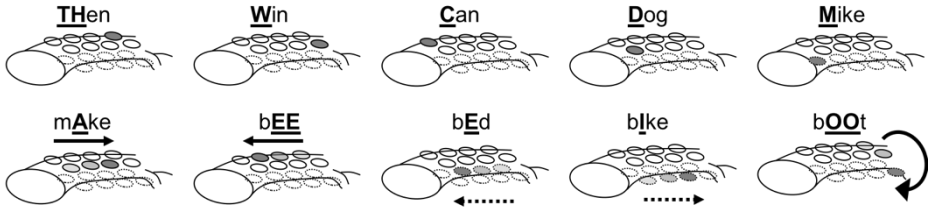

Fig. 2. Custom haptic patterns and associated phonemes used for training and testing.

2.3 Evaluation Criteria

We evaluated the quality of the algorithms according to discriminability, learnability, and generalizability. 'Discriminability' refers to the ability to detect differences in haptic patterns that represent phonetically different words, and to recognize as identical haptic patterns that represent phonetically identical words. Evaluation of discriminability is rapid since learning is not required. In contrast, 'learnability' allows the correct identification of a word, represented by a haptic pattern, some period (minutes-days) after exposure. Finally, 'generalizability' allows the identification of novel patterns by drawing on learned associations between components of the haptic pattern and lexical information. By comparing performance on a subset of phonemes and words, we can gain insight into the quality of the coding algorithms without requiring the arduous process of first learning an entire vocabulary.

2.4 Study 1: Discriminability

The 3 acoustic algorithms were compared in terms of *sensitivity* (fraction of pairs of different words that are correctly identified as different) and *specificity* (fraction of pairs of identical words that are correctly identified as not different). The PH algorithm was not included because its custom patterns were thought to be sufficiently discriminable by design. Four 8-actuator displays were used with actuators placed in a linear configuration on the dorsal forearms and lateral upper arms, as shown in Fig. 1.

Custom recordings from a single male native speaker of American English were used to form test pairs of monosyllabic words that differ in a single phoneme: beat/but, bet/debt, bet/met, bet/pet, bet/vet, bird/bud, book/bought, bought/but, boy/buy, fat/vat, heat/hit, lit/wit, met/net, pet/tet, sat/shat, and that/vat. These were selected to 1) give good representation of the different phonemic groups, 2) provide comparison phonemes that are similar in production and hence perceptually close (e.g., 'b' and 'd' in bet/debt are both plosives that are produced toward the front of the mouth), and 3) minimize the number of distinct phonemes. Negative (beat/beat) and positive (boy/sat; differs in all phonemes) controls were included. Haptic patterns lasted 200-300 ms with an inter-word interval of 350 ms. A 500 ms delay was imposed between response and subsequent trial. For each of the 3 algorithms, the 18 pairs were repeated 6 times for a total of 324 trials, with the participant tasked with indicating whether the pairs were "same" or "different". The order of presentation of each pair was randomized over all repetitions and algorithms, as was the word order within pairs.

The null hypothesis, that responses associated with different algorithms are governed by the same underlying statistics, was evaluated with permutation testing, which, advantageously, makes no assumptions about the properties of the data [12]. Briefly, the response data was randomly relabeled and the test statistic (difference in sensitivity) was recalculated based on the new labeling. The process was repeated 50,000 times, and a distribution of the permutation test statistic was generated. The probability that the actual test statistic arose from this distribution was estimated. If the probability was less than 0.0167 (p=0.05 with Bonferroni correction for 3 comparisons) the null hypothesis was rejected. An ANOVA with the algorithms as within-subject factors was a pre-specified secondary analysis.

2.5 Study 2: Learning

The learnability and generalizability of haptic patterns encoded from single words were tested using the PH and DSP algorithms (results of Study 1 suggested DSP yielded the best discriminability). 32-channel actuators arranged linearly over both arms were used for DSP. PH used two 8-channel displays arranged as a 4x4 grid on the left forearm, with 2x4 grids on the dorsal and volar surfaces.

Training for a practical TLC system would, ideally, be rapid, engaging, and fun. The potential for implicit learning in the context of video games has previously been demonstrated [13]. In addition, the ability of video games to engage the user is well known [14]. We therefore designed a game-based training paradigm with the goals of incorporating implicit, explicit, incidental, and adaptive learning approaches; and to

reinforce haptic-audio cross-modal association. The implementation of the game, called RoboRecycle, was performed by Coatsink Software (Sunderland, UK) using the Unity game engine (Unity Technologies SF, San Francisco) with haptic controls stored as .wav files and delivered through a USB audio interface using Max 7 (Cycling '74, San Francisco).

RoboRecycle had 3 modes (Fig. 3). In *Explore* mode, the user initiated haptic playback by selecting from a menu of words to gain familiarity with the stimuli that would be delivered during gameplay, and to begin to form associations between words and haptic patterns. In *Game* mode, the user attempted to rescue robots before they were consumed by a recycler. A haptic pattern was repeatedly delivered as a conveyor belt carried the robot toward its fate. The player was cued with the associated spoken word, which was delayed to avoid cross-modal masking. As levels progressed, the number of cued words decreased, so at the highest level there were no spoken cues. The player selected an answer from a menu. If correct, the robot was rescued. If incorrect, or if the robot reached the end of the conveyor belt, it fell to its doom. In both cases the correct answer was shown. The player progressed to the next level if > 70% of responses were correct. If < 50% were correct then the player dropped back to the previous level. If accuracy was between 50 and 70% then the player repeated the level. In *Test* mode, the player received a single presentation of each haptic pattern, and selected an answer from a menu of possibilities. No feedback was provided.

Fig. 3. Screenshots from RoboRecyle Explore, Game, and Test modes

24 words with 2-3 phonemes/word were used in 3 sets as follows: List 0: ace, aid, same, sue; List A: cake, die, make, mime, mood, me, they, weed, wide, woo; List B: deed, dime, do, doom, key, may, meek, wake, why, womb. Words were selected to be representative of the main features of articulation. Lists A and B were derived from the 10 distinct phonemes shown in Fig. 2. Recordings were made by a single male native speaker of American English. List 0 was for demonstration only. The order in which Lists A and B were used in the study was counterbalanced among participants.

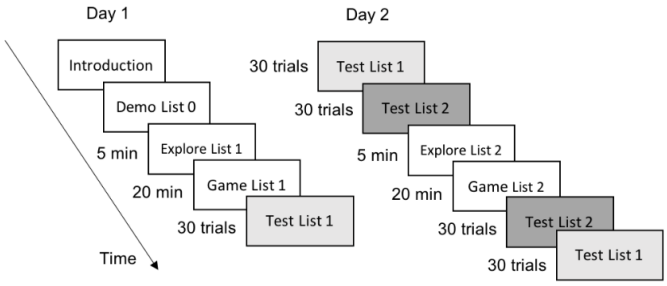

Fig. 4. Learning study design

The study design is illustrated in Fig. 4. The study began with a demonstration of the RoboRecycle game using List 0. This consisted of a brief guided tour of each game mode with the participant able to activate controls and experience the haptic stimuli.

Training occurred in 2 phases. In the first, the participant was allowed 5 minutes to initiate haptic playback using Explore mode with List A or B ("List 1"). During this phase, participants in the PH algorithm group had access to a visual representation of patterns associated with each phoneme. Participants in the DSP group received the following description of coding rule: "The algorithm maps frequencies in the speech signal to locations on the arm. Frequencies increase progressively up the arms, with lowest mapped to the left wrist, low-medium to the left shoulder, high-medium to the right wrist, and highest to the right shoulder." The second training phase used Game mode, which was stopped after 20 min or when the participant reached at least 90% correct at the highest level.

The participant was tested immediately after completing the Game phase. Each of the 10 haptic patterns was presented 3 times, for a total of 30 trials in random order. At each trial, the haptic pattern was delivered once, and the participant selected an answer from a menu of the 10 words. No feedback was provided.

The second day began with testing the 10 haptic patterns from the previous day (List 1) to evaluate retention. This was followed by a test of 10 new haptic patterns (List 2) to test generalization. The testing protocol was identical to that used on the first day. After testing, the participant again underwent training using RoboRecycle's Explore and Game modes with the previously unused word list (List 2), using the same protocol as Day 1. After the training, the participant was tested on the 10 current words first (List 2), and then on the words that were learned the previous day (List 1).

To evaluate learnability and generalizability, mixed 2x5 ANOVA was conducted on test accuracy with the 2 algorithms and 5 time points (post-training of List 1 on Day 1; and retention of List 1, pre-training of List 2, post-training of List 2, and 2nd retention of List 1 on Day 2) as within-subject factors, followed by post hoc analyses on any significant effects with Bonferroni adjustment. In addition, the test accuracy at 5 different points was compared against 10% chance level using one-sample *t*-tests with Bonferroni adjustments for multiple comparisons. ANOVA rather than permutation testing was used because of its ability to support multiple comparisons across multiple dimensions. Moreover, the rates of level progression during Game (training) mode

across time for different algorithms and on different days were estimated using linear regression.

3 Results

3.1 Study 1: Discriminability

10 participants were recruited from our group after providing informed consent (2 female; ages 34.9±9.2, range 24-54). Previous haptic experience varied from limited to extensive.

Detection sensitivity for FD, DSP, and AE was 0.58±0.20, 0.60±0.19, and 0.38±0.23, respectively (ave±SD) (Fig. 5). Permutation testing revealed no significant difference between FD and DSP (p=0.37), but the differences between FD and AE (p<0.001) and between DSP and AE (p<0.001) were significant with Bonferroni correction. These findings were confirmed with ANOVA, which showed no significant differences for specificity (p=0.87) but did for sensitivity (p<0.01) as a result of significantly lower sensitivity of AE than both FD and DSP (p=0.02).

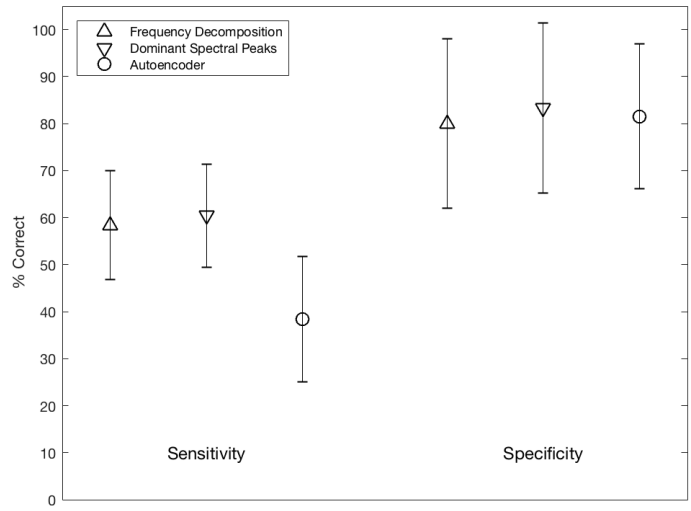

Fig. 5. Sensitivity and specificity (ave±SEM) for FD, DSP and AE. Specificity was not significantly different among the algorithms, but AE yielded significantly worse sensitivity.

3.2 Study 2: Learning

16 participants without previous haptic experience were enrolled in the learning study after providing informed consent, with 8 randomized to the DSP group (age 31.3±12.3, range 23-57; 4 females), and 8 to the PH group (age 28.8±7.3, range 22-44; 2 females).

Test performance is shown in Fig. 6. Mixed ANOVA analysis revealed a significant main effect of algorithms (p=0.01), with omnibus test accuracy higher in PH than DSP

group. There was a significant main effects of time points (p<.001), as well as time points by algorithms interaction (p=0.01).

Post-hoc analyses showed that the PH group significantly outperformed the DSP group at pre-training of List 2 (p<0.01), which was also the only time that a group's test accuracy (DSP) was not significantly different from 10% chance level (PH: p = 0.01, 0.01, 0.03, 0.025, and < 0.001; DSP: p=0.01, <0.001, 1.0, 0.01, and 0.01; for post 1 on Day 1, and retention 1, pre 2, post 2, and 2nd retention 1 on Day 2, respectively). PH performance was also significantly better on post-training of List 2 (p=0.01) and the 2nd retention test of List 1 on Day 2 (ps<0.01). There were no significant differences among the 3 tests of List 1 for either group (p>0.05), indicating that a decrease (or increase) in retention with the passage of time and acquisition of new material could not be detected. Critically, the recall test result immediately after training was significantly better in List 2 than List 1 for the PH group (p=0.01), but not for the DSP group (p=1.0), suggesting better generalizability in the PH group. There was no significant difference in test performance on List A vs B for either algorithm at any test point.

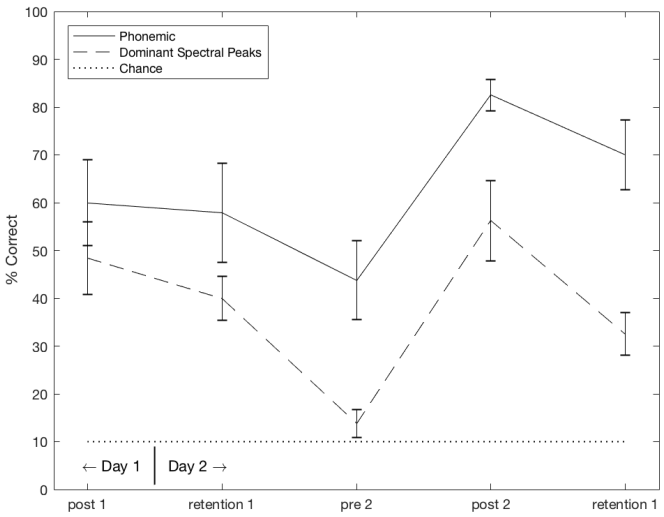

Fig. 6. Test performance (ave±SEM) for PH (solid), for DSP (dashed), and expected from chance (dotted).

The rate of level progression was not significantly different between PH and DSP groups on Day 1 (p=0.17), but was significantly faster for PH on Day 2 both when compared to DSP on Day 2 (p<0.001), and when compared to itself on Day 1 (p<0.001), indicating improved learnability with the PH algorithm with additional exposure. In contrast, rate of progression was slower for DSP on Day 2 compared to Day 1 (p=0.01), suggesting additional exposure interfered with learning.

Combining word lists, the PH group demonstrated 76.3±14.7% (ave±SD) accuracy on the 20 words after 50 min of focused training (excluding learning effects of introduction and test taking). For DSP accuracy was 44.4±16.2%.

4 Discussion

This work presents 4 candidate algorithms for translating natural language to haptic stimuli for transcutaneous communication, along with evaluation methodology based on discriminability, learnability, and generalizability.

The three acoustic algorithms we tested were not significantly different in specificity, however, the ability to detect single phoneme differences in word pairs was significantly worse using AE compared to both FD and DSP. Using a set of 20 words composed of 10 phonemes, learning was possible with both DSP and PH, but participants using the PH algorithm showed significantly better recall, and demonstrated an ability to generalize learned phoneme associations to new words. Overall, the PH group demonstrated 76.3±14.7 accuracy on 20 haptic words after 50 min of focused training.

The AE algorithm represents an instantiation of a well-known class of unsupervised machine learning algorithms that has demonstrated success with data compression. However, the encodings are not readily interpretable by humans due to the high entropy that results from elimination of redundancy during compression. We sought to mitigate this by imposing constraints during training that would facilitate interpretation of haptic stimuli. Despite this, performance in discrimination was inferior to frequency-to-spatial mapping algorithms. Study participants described the AE stimuli as an unstructured, random buzzing, which suggests that entropy remained unacceptably high. More aggressive implementation of constraints was subsequently explored, including adding temporal stability (infrequent changes in state are preferred) and spatial stability (spatial correlation among nearby actuators is preferred). However, reduced entropy came at the expense of lower information capacity to an extent that precluded an acceptable compromise between high entropy and low information content. Further improvement may be possible by incorporating a more refined psychophysical model in network training based on user testing.

Phoneme-based algorithms have several attractive features. Discriminability is guaranteed by design, and mnemonic aids can be incorporated, such as mechanics of articulation [10] or qualitative features of the sound [3]. Furthermore, the language is broken down into a relatively small number of fundamental units (building blocks) with which speakers are intimately familiar, and which form a basis of spoken language acquisition. At the most basic level, once the phoneme/haptic symbol association is learned, comprehension is guaranteed provided the delivery rate is sufficiently slow. Analogous to natural language acquisition, we expect decreased cognitive load along with increased fluency and delivery rate as users master phonetic components and start to recognize longer lexical structures in a ‘chunking’ process [15]. In contrast, lack of generalization with the DSP algorithm and degradation of performance when new words were added suggest a rote learning process in which acoustic mappings, while discriminable, lacked sufficient perceptual or semantic congruency to the lexical information.

Discriminability, learnability, and generalizability are useful criteria for rapid and effective initial screening of candidate algorithms. Regarding discriminability of word pairs: the algorithm, the haptic display, and the participant together comprise a detection test, which makes analysis amenable to the tools of signal detection theory [16]. In

this context, an earlier version of the protocol called for a continuous graded response to reflect the participant's perception of the degree to which the two stimuli differed. However, several drawbacks became apparent: interpretation of the task varied greatly among participants, cognitive demand was greater resulting in longer response times and greater participant fatigue, and while responses were consistent when word pairs were identical or very different, intermediate pairs yielded inconsistent, highly variable responses. For these reasons we adopted the two-alternative, forced-choice protocol presented here.

Viewing the algorithm, haptic display, and participant as an integrated unit also highlights the challenge of evaluating an individual component of the triad. A display that does not reliably deliver the intended stimulus intensity (for example, due to excessive sensitivity to backing pressure) will make it more difficult to detect differences in algorithm performance. Similarly, poor participant motivation, including skepticism of the premise that transcutaneous communication is possible, can mask differences in algorithms or displays. The location for the display was chosen primarily for convenience. The two halves of the display covered both glabrous and hairy skin, and included regions near and far from anchors (wrist, elbow). Different performance can be expected using a display at locations that have greater or lesser sensitivity.

Learnability and generalizability imply discriminability. They more directly reflect the target task but are more demanding and time consuming to test. We minimized these drawbacks by testing a subset of phonemes with a small number of words. Despite the small word set, statistically significant differences in algorithm performance were apparent.

There are several limitations to this work. We take as a premise that discriminability, learnability, and generalizability, as defined here, reveal features of algorithm performance that predict success in the target use case, but this has not been validated. Candidate algorithms were tested on small subsets of phonemes and vocabularies with a modest number of participants, and learning and retention were tested over relatively short time scales. The criteria were applied to isolated words without context, rather than phrases or sentences as would be used in a practical system, and which might aid in interpretation of individual words. The tested algorithms represent specific implementations of their classes; better performance may be possible with further parameter optimization. Future studies will seek to address these limitations.

5 Conclusion

Discriminability, learnability, and generalizability are useful criteria for the efficient evaluation of candidate language-to-haptic translation algorithms. The acoustic algorithms we tested did not differ significantly in specificity, but AE was significantly worse than FD and DSP in sensitivity. Individual words could be learned with both the PH and DSP algorithms, but for PH test performance was significantly better and there was evidence of an ability to generalize learning to previously unseen words. Participants in the PH group demonstrated 76.3±14.7% (ave±SD) correct on 20 haptic words after 50 min of focused training.

Acknowledgements
This work was supported by Facebook, Inc.

References

1. Reed, C.M., Rabinowitz, W.M., Durlach, N.I., Braida, L.D., Conway-Fithian, S., Schultz, M.C.: Research on the Tadoma method of speech communication. J. Acoust. Soc. Am. **77**, 247–57 (1985). doi:10.1121/1.392266
2. Brooks, P.L., Frost, B.J., Mason, J.L., Gibson, D.M.: Continuing evaluation of the Queen's University tactile vocoder II: Identification of open set sentences and tracking narrative. J. Rehabil. Res. Dev. **23**, 129–138 (1986)
3. Israr, A., Meckl, P.H., Reed, C.M., Tan, H.Z.: Controller design and consonantal contrast coding using a multi-finger tactual display. J. Acoust. Soc. Am. **125**, 3925–35 (2009). doi:10.1121/1.3124771
4. Galvin, K.L., Mavrias, G., Moore, A., Cowan, R.S.C., Blamey, P.J., Clark, G.M.: A comparison of Tactaid II+ and Tactaid 7 use by adults with a profound hearing impairment. Ear Hear. **20**, 471–482 (1999). doi:10.1097/00003446-199912000-00003
5. Oller, D.K.: Tactile AIDS for the hearing impaired: An overview **16**, 289-295 (1995). doi: 10.1055/s-0028-1083726
6. Miyamoto, R.T., Robbins, A.M., Osberger, M.J., Todd, S.L., Riley, A.I., Kirk, K.I.: Comparison of multichannel tactile AIDS and multichannel cochlear implants in children with profound hearing impairments. Am. J. Otol. **16**, 8–13 (1995). doi:10.1016/0165-5876(95)97416-6
7. Brooks, P.L., Frost, B.J.: Evaluation of a Tactile Vocoder for Word Recognition. J. Acoust. Soc. Am. **74**, 34–39 (1983). doi:10.1121/1.389685
8. Muda, L., Begam, M., Elamvazuthi, I.: Voice Recognition Algorithms using Mel Frequency Cepstral Coefficient (MFCC) and Dynamic Time Warping (DTW) Techniques. **2**, 138–143 (2010). doi:10.5815/ijigsp.2016.09.03
9. Hinton, G.E., Salakhutdinov, R.R.: Reducing the dimensionality of data with neural networks. Science (80-.). **313**, 504–507 (2006). doi:10.1126/science.1127647
10. Zhao, S., Israr, A., Lau, F., Abnousi, F.: Coding Tactile Symbols for Phonemic Communication. Proc. CHI'18 (2018). doi:10.1145/3173574.3173966
11. Israr, A., Poupyrev, I.: Tactile Brush : Drawing on Skin with a Tactile Grid Display. Proc. CHI'11. 2019–2028 (2011). doi:10.1145/1978942.1979235
12. Pesarin, F., Salmaso, L.: The permutation testing approach: a review. Statistica. **70**, 481–509 (2010). doi:10.6092/issn.1973-2201/3599
13. Lim, S. joo, Holt, L.L.: Learning Foreign Sounds in an Alien World: Videogame Training Improves Non-Native Speech Categorization. Cogn. Sci. **35**, 1390–1405 (2011). doi:10.1111/j.1551-6709.2011.01192.x
14. Przybylski, A.K., Rigby, C.S., Ryan, R.M.: A Motivational Model of Video Game Engagement. Rev. Gen. Psychol. **14**, 154–166 (2010). doi:10.1037/a0019440
15. Hewlett, D., Cohen, P.: Word segmentation as general chunking. Proc. Fifteenth Conf. Comput. Nat. Lang. Learn. 39–47 (2011)
16. MacMillan, N.A.: Signal Detection Theory. In: Stevens' Handbook of Experimental Psychology (2002)