# Conversing Using WhatsHap: a Phoneme Based Vibrotactile Messaging Platform

David Marino[1,3] Maurício Fontana de Vargas[2,3] Antoine Weill--Duflos[1,3] Jeremy R. Cooperstock[1,3]

*Abstract*— **We demonstrate the feasibility and experience of having a haptic conversation using WhatsHap: an instant messaging system that delivers speech or text as a sequence of vibrotactile representations of English phonemes to the arm. Previous haptic speech communication studies established feasibility in single-phoneme or word-level encodings, but did not investigate how such communication functions in practice with real-time remote conversation between two individuals. Participants used WhatsHap through the framework of a joint communication task, where they had to converse to achieve a goal, with 88% of all tasks successfully completed. We analyze conversations and user interviews both qualitatively and quantitatively, describing considerations when building a system to mediate conversation haptically, exploring influences on user conversational experience, and offering an account of how linguistic structure changes to accommodate such a mode of communication. In this regard, phoneme-based haptic conversation led to linguistic forms distinct from written and spoken English. Additionally, participants felt that haptic conversation was best suited for information-centered communication in contexts where there is shared knowledge between users.**

## I. INTRODUCTION AND RELATED WORK

Natural language is inherently multimodal, though when transmitted remotely it is frequently received unimodally through audition or vision. For example, a text message transmits language visually. There are, however, many circumstances where both audition and vision may be infeasible. For example, reading a text message while walking near a busy intersection can prove dangerous if visual and auditory senses are preoccupied. Similarly, airline pilots [1], emergency care unit [2] and power plant workers [3], may reach visual and auditory sensory saturation. It is thus useful to be able to encode natural speech in alternative modalities, such as touch, to offload visual or auditory saturation.

The best-known example of using the skin as a communication channel is the Tadoma method, in which deaf-blind users receive speech by placing their hand on the talker's face. Additionally, research going back as far as 1928 explored the use of vocoders [4]. With this approach, the speech signal is processed in real time using a bank of band-pass filters to deliver temporal, intensity, or spectral information on a tactile display consisting of a vibrotactile transducer for each filter band. Hearing subjects trained for 55 hours a chieved 80% accuracy on a 150-word set [5], while deaf participants trained for 235 hours (spread across 47 weeks) achieved 90% on a 135-word list [6]. However, experiments with commercially available vocoders demonstrated that they could not be used for understanding speech without the support of lipreading [7]. Recent effort on haptic communication has focused on rendering English words as a sequence of small discrete units such as letters [8], [9], [10], Morse code [10], or English phonemes [11], [12], [13], [14], [15], [16], assuming speech-to-text as a first layer in the system.

Delivering a message based on its phonology offers the advantages of disambiguation and efficiency. For languages such as English with a deep orthography, one grapheme can map to many phonemes For example, a $\langle c \rangle$ may be an /s/ as in $\langle \mathbf{c}icada \rangle$ or a /k/ as in $\langle \mathbf{c}at \rangle$. In terms of efficiency, multiple graphemes may be required to represent a single phoneme. For example, in the word $\langle morale \rangle$, the -ale maps to /æl/ with the $\langle e \rangle$ unpronounced. This makes an English orthography-based delivery system slower [15] and less transparent. Zhao et al. utilized a 6-channel vibrotactile display to show that rendering haptic symbols representative of English phonemes based on articulatory features bolstered accuracy levels in word identification tasks [15]. Tan et al. demonstrated that users were able to acquire a vocabulary of 500 words using a $4{\times}6$ tactor array [16]. Using MISSIVE [12], a multi-sensory (radial squeeze, lateral skin stretch, vibration) device worn on the upper arm, participants achieved 87% word identification accuracy after 100 min of training on a set with 150 words. Similarly, de Vargas et al. [14] reported an accuracy of 94% under the same training and testing protocol and 45% with an open-answer format using two voice-coil actuators.

Regardless of the mapping strategy, the aforementioned researchers were able to demonstrate the feasibility of delivering speech through the sense of touch after non-extensive training. However, these studies ultimately focus on vocabulary learning, using unidirectional communication, with controlled and unnatural speech delivered by a training software, which would randomly select the vocabulary and translate into the haptic stimuli. For the vocoder approach, researchers have explored the rendering of entire sentences [17], [18], but their devices were intended to function as a complement to lip-reading, for which participants were also presented with video recording of the speaker's face. On discrete-mapping systems, de Vargas et al. [19] has investigated the delivery of phrases with untrained vocabulary, but no prior work has studied nor demonstrated effectiveness of complex conversation between two interlocutors in which at least one person receives all communication haptically.

[1] Department of Electrical and Computer Engineering, McGill University, Montréal, Canada

[2] School of Information Studies, McGill University, Montréal, Canada

[3] Centre for Interdisciplinary Research in Music Media and Technology, Montreal, Canada

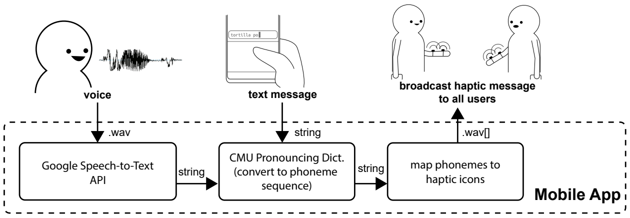

voice
text message
broadcast haptic message to all users
.wav
string
.wav[]
Google Speech-to-Text API
string
CMU Pronouncing Dict. (convert to phoneme sequence)
string
map phonemes to haptic icons
Mobile App

Fig. 1: WhatsHap's speech/text-to-haptics algorithm.

In this work, we explore how such "speech-to-haptic" devices would work in practice when users are tasked with composing and receiving phrases haptically during live conversation. We gain deeper individual insights of the user's experience in communicating haptically by conducting a user study where trained users engaged with conversation partners to achieve a certain goal, receiving all speech as haptic stimuli through WhatsHap—a mobile, wearable, alternative communication device capable of speech replacement through dual-channel vibrotactile representations of discrete English phonemes.

## II. System Design

### A. Overview

The system consists of a vibrotactile device connected to a smartphone running a messaging app. The vibrotactile device comprises of two voice-coil based actuators (Haptuator Original Tactile Labs, Montreal, Model no. TL002-14-A) [20], secured with armbands on the dorsal side of the user's forearm. One actuator is placed close to their wrist, and the other close to their elbow. The messaging app runs in a web browser (Figure 1), and allows the user to send a haptic encoding of their speech or text message.

In the case of speech, the app uses Google Speech-to-Text API[1] to convert the user's utterances into text. It then uses the CMU Pronouncing Dictionary[2] to obtain a phonemic representation of the utterance's words in North American English. The user can then broadcast their utterance as a stream of haptic symbols to all other interlocutors that are also using the system. If using text input, the algorithm follows the same steps excluding speech-to-text conversion. Haptic-encoded speech is then delivered with 1 s inter-phoneme intervals and 3 s inter-word intervals.

### B. Phoneme-to-Haptic Mapping

Our rendering strategy used on WhatsHap supports 24 English phonemes—15 consonants (/p, b, t, d, k, g, f, v, ð, s, z, m, n, l, ɹ/), 5 vowels (/i, ɛ, ʌ, u, ɑ/), and 4 diphtongs (/eɪ, ɑɪ, ɑʊ, oʊ/) chosen based on frequency of use during casual conversation [21], and distinction between other phonemes such that similar-sounding phonemes could be rendered with the same haptic symbol. With ten substitutions ( æ→ɛ, j→i, ɪ→i, ɝ→ʌɹ, ʊ→u, ɔ→ɑ, w→u, ɳ→ng, ɔɪ→ɑɪ, θ→f), WhatsHap is able to convey a larger set of words while keeping the minimal hardware design and reducing the training time needed for learning all haptic phonemes covered. The drawback is that minimal pairs, i.e., words differing by only one phoneme (e.g., man–men, eat–it, peel–pill) involving such potentially substitute phonemes can only be correctly identified with the support of the other words in the phrase or the communication context previously established (e.g., talking about food).

The haptic stimuli are designed to provide a natural mapping between the haptic sensation and the phoneme's associated anatomical physicality. The principles are summarized as follows, while an in-depth explanation can be found in reference [14]:

*1) Consonants:* Audio of the isolated consonants was obtained from recordings of a native English speaker.[3] The raw audio signal was processed to enhance characteristics inherent to the *manner* of articulation, i.e., how speech organs modulate the air flow to produce the phoneme. For example, the high-frequency and turbulent sound of fricative phonemes (/f, v, s, z/) caused by the air going through a narrow gap between the lips or teeth was emphasized by a high-pass filter, while the strong and short puff of air characteristic of the plosives (/p, t, k/) was enhanced by changes in gain. The duration of the consonant stimuli is also determined by the manner of articulation, with plosives being the shortest (20-35 ms) and nasals the longest (550 ms). Finally, the inter-tactor intensity difference (IID) of the two channels was adjusted to create a spatial panning indicative of the phoneme's *place* of articulation on the vocal tract: phonemes produced in the front of the mouth (e.g., /b/) are mapped towards the distal region of the forearm, while phonemes produced in the back of the oral cavity (e.g., /g/) are rendered in the proximal region.

*2) Vowels and Diphtongs:* Audio of the individual vowels were synthesized in Praat [22] for better control over length and the fundamental frequency (F0) in comparison to audio recordings. All vowels are 750 ms in length and were synthesized with a F0 of 140 Hz. The formant frequencies (F1 and F2) used in the syntheses were provided by the software tools, based on real speech average values. Since vowels do not have distinct features in terms of manner of articulation, the only enhancement performed was for creating the spatial panning indicative of the phoneme's *place* of articulation. To improve discrimination between vowels and consonants, a unique fade-in and fade-out effect was employed in all vowels. The articulation movement characteristic of diphthongs is reproduced by linearly varying the IID between the values of the vowels composing the diphthong, resulting in the perceptual illusion that the vibration moves between the starting and ending vowels' locations. All diphthong stimuli are 1.5 s in length.

## III. User Study

### A. Participants

We recruited 5 participants to play two different roles: *haptic listeners*, who received messages from a conversation

[1] https://cloud.google.com/speech-to-text/docs/

[2] http://www.speech.cs.cmu.edu/cgi-bin/cmudict

[3] www.jbdowse.com/ipa

partner (CP) in the form of haptic phonemes and responded via text, and *speakers*, who could either speak or type into the system, and received messages only by text. We distinguished between these two roles to allow us to study how naive users who had no prior haptic phoneme training would learn to use this system to converse with haptic listeners. We use H# to refer to a particular haptic listener, and S# to refer to a particular speaker.

We recruited Haptic Listeners from a previous study who were able to identify at least 65% of words rendered within phrases using the same haptic encoding. Three participants who achieved accuracy scores of 91%, 85%, and 68% after training for 250 min were recruited. The last time they had used the device was approximately five months before this study.

We recruited 2 Speakers with a speech science background who had never encountered any version of the system before.

All participants provided informed consent of the experiment protocol, following Research Ethics Board guidelines, and received compensation of CAD $20.

### B. *Procedure*

The experiment was held in a laboratory setting. *Haptic listeners* were in a room that was acoustically and visually isolated from the *speakers*, and thus all communication was conveyed through WhatsHap.

Haptic listeners wore headphones playing masking pink noise, with the haptic apparatus attached to their right arm.

At the beginning of each session, haptic listeners were given a 50 min review session to refresh themselves on the haptic encoding. The review consisted of unstructured self-training activities on individual phonemes, words, and phrases, delivered through a computer. Then, during the main phase of the study, both users were asked to converse using WhatsHap through the framework of a joint communication task (detailed below).

Each of the experts completed two sessions, each involving a separate communication task with a different speaker. The speakers returned for successive sessions with different listeners for a total of three sessions. During a session, turn length, message replays, accuracy scores, task performance, and message complexity were measured.

At the end of each session, we conducted a semi-structured interview with participants to gather their experiences. We transcribed this data and conducted content analysis on the transcripts. The first two authors performed separate analyses using an inductive open coding, developing the emerged codes into themes and subthemes through axial coding. Then, they met twice to discuss and reflect on their codes, identifying similarities in the emerged themes. Finally, the first author combined these analyses to come up with a single set of codes, themes, and subthemes.

*1) Description of communication tasks:* We created two communication activities based on the concept of task-based language learning [23], an approach extensively used in second-language instruction. These tasks focus on achieving a certain goal, rather than a pure linguistic outcome, and have a gap (information, reasoning, or opinion [24]) to be overcome by the CP. Participants were free to use whatever words they wanted to accomplish the tasks. A task was "successfully completed" if the gap was overcome.

In the first task (information-gap), the haptic listener played the role of a chef and received a timetable containing a cooking plan for the week, including missing ingredients for each recipe. The speaker's tasks were to find out: what is on the menu, what ingredients (and quantity) are needed, and hours of operation to supply it.

In the second task (reasoning-gap), the speaker was requested to invite their CP to perform any activity they would like, e.g., play golf, and to schedule it at a time both were available according to the timetables they received. Approximately half of the time slots in the timetables were occupied to increase the likelihood of negotiation between participants.

## IV. Results and Discussion

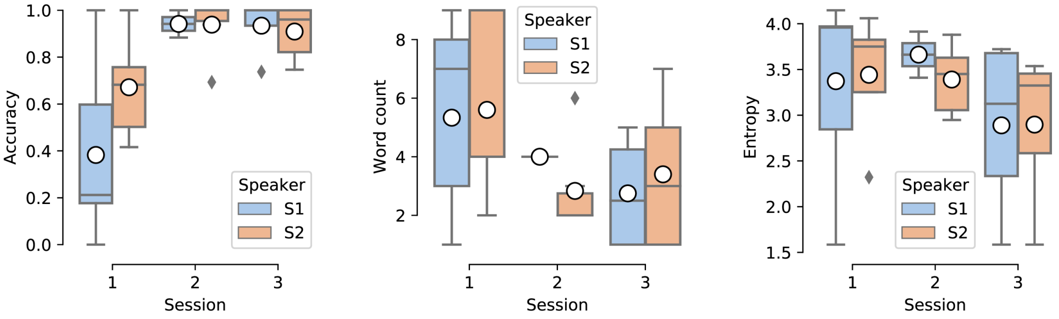
We first present a quantitative description of how the haptic conversations evolved through the study sessions. Then, we report our qualitative findings regarding users' experiences communicating through the system.

### A. *Quantitative Description of Haptic Conversations*

Response time (RT) between CPs tended to be longer than natural communication, with each turn lasting an average of 2.84 min ($SD_{time} = 1.90\,\text{min}$). RT was measured as the time between when the haptic stimuli finished playing to when the haptic speaker sent their message ($SD_{replay} = 2.01$). Listener accuracy was calculated as a function of how much phonological content was successfully understood, according to the phonological edit distance [25]: haptic listeners were asked to write down the content of the messages they received haptically. We then compared the haptic listener's reiterated message with the original message by calculating the Levenshtein edit distance (i.e., the number of operations required to transform one string to another) weighted by the distinctive features [26] associated with each phoneme. We normalized the edit distance as a value between 0 and 1 expressed as in Equation 1. In the normalized phonological edit distance, 1 = a perfect match, and 0 = a complete mismatch.

$$1 - \min\left\{\frac{phonoEditDistance(str_1, str_2)}{phonoEditDistance(str_1, \text{“”})}, 1\right\} \qquad (1)$$

Haptic listeners exhibited average accuracy scores of 0.73, with a clear upward trend in accuracy over time (Fig. 2a). Regardless of individual phoneme accuracy, participants as a whole were able to comprehend the essence of what their conversation partner was saying, as 87.5 % of all communication tasks were successfully completed. In the first session, to fully comprehend the speaker's intentions, haptic listeners replayed messages a mean of 3.36 times, with a maximum of 8 repetitions for a single message. By the final session, listeners only replayed messages 1.88 times on average, with a maximum of three replays for a single message.



(a) Accuracy based on the normalized edit distance for the reiterated phrase compared to that of the naive user (speaker)

(b) Word count

(c) Information entropy

Fig. 2: Boxplots of the evolution of messages sent by naive speakers over the three sessions. White dots indicate the mean.

TABLE I: A high level overview of the categories and themes that emerged from content analysis of interview data

| Theme | Categories |
|---|---|
| Context and linguistic structure for effective communication | Adjusting communication style to task<br>Suprasegmental aspects lost in haptic communication<br>Type of communication the system is best suited for |
| Influences on cognitive load | Training<br>Mental state and performance<br>Encoding system adjustments for reduced CL<br>Linguistic and cultural background |

As users learnt more about the system, the number of words per message, along with the average information entropy [27], decreased over time, as seen in Figures 2b and 2c. There was also a reduction in variance in both of these variables. The decrease in entropy exhibited an inverse relationship with accuracy scores.

During the first session, S1 tended to use long sentences ($\bar{X}_{words} = 5.33$, $SD_{words} = 3.19$, $\bar{X}_{entropy} = 3.37$ bits). Their first strategy to improve comprehension was to speak slower and clearer into the microphone, with minimal changes to the words or structure of their sentences. This strategy was sub-optimal because the system delivers haptic phonemes to users at constant intervals regardless of the suprasegmental characteristics of the input speech. Later, they opted to use a distinct form of syntactic abbreviations. This is further elaborated upon in our qualitative findings. By the final session, S1's utterances had reduced to an average of 2.75 words per message ($SD_{words} = 1.78$, $\bar{X}_{entropy} = 2.88$ bits).

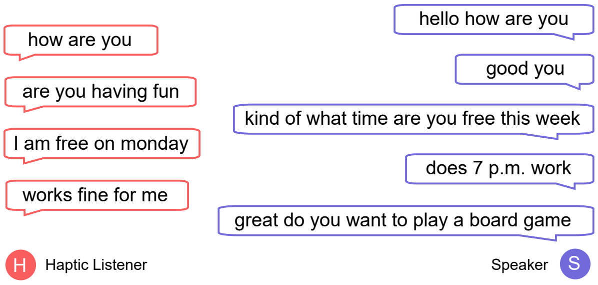
### B. *Qualitative Findings and Observations*

Content analysis yielded 2 key themes: *context and linguistic structure for effective communication*, and *influences on cognitive load*. These themes along with their component categories are outlined in Table I.

#### *1) **Theme 1: Context and linguistic structure for effective communication**:*

[*Category 1A: adjusting communication style to task*] — When S1 and S2 initially used the system, they spoke very similarly to everyday conversation, with many phatic expressions (Fig. 3) One session later, their utterances became more direct, with pleasantries removed (Fig. 4). They tended to minimize function words (e.g., prepositions, determiners) from their utterances, sticking mainly to content words (e.g., nouns, verbs). S2 noted, "pronouns and function stuff. . . I would try and get rid of those and be like 'Wednesday available'." S1 and S2 in later sessions tended to drop the subject of their sentences. H3 noticed a change in communication style in their own session, commenting, "I think the messages were more to the point. More direct. Instead of full sentences, it was like 'this is what I need'. It made it easier to interpret." A unique way of shortening haptic messages emerged from conversation, distinct from everyday texting and speech. S2: "If I was to send this [haptic message] as a text message and not speak it I wouldn't say the full day of the week, like I would just put 'mon' for Monday, but then it's weird to say [out loud] 'mon'." H3 said that if they were to receive a haptic message that included an orthographic or phonetic abbreviation such as "mon" for "Monday", it would lead to confusion. S1 and S2 also did not use shortenings commonly seen in spoken language such as contractions like "what is" → "what's", opting to stick to the full form of words.

[*Category 1B: suprasegmental aspects lost in haptic communication*] — Though phoneme-based haptic speech rendering has shown promise over vocoder-based approaches in terms of raw haptic listener accuracy scores without needing to see the speaker's face, a purely phonemic approach masks vocal prosody and discards many *suprasegmental* aspects of speech such as tone, stress, and rhythm, among others. These features serve a variety of functions in understanding speech: from inferring emotions and humor, to lexical semantics and phonemic distinctions. One function of prosody in English is in distinguishing yes-no questions from statements. S1 mentioned "prosody would be helpful in questions. There



hello how are you

how are you

good you

are you having fun

kind of what time are you free this week

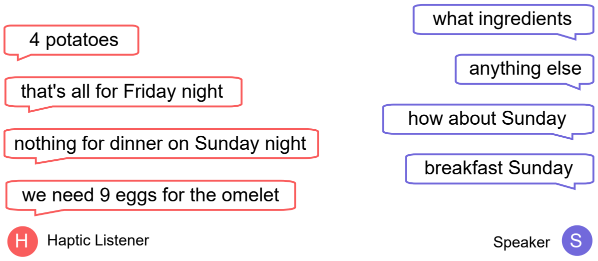
I am free on monday

does 7 p.m. work

works fine for me

great do you want to play a board game

Fig. 3: Transcript of S2 conversing in the first session



what ingredients

4 potatoes

anything else

that's all for Friday night

how about Sunday

nothing for dinner on Sunday night

breakfast Sunday

we need 9 eggs for the omelet

Fig. 4: Transcript of S2 conversing in their second session

could be like a statement versus question prosodic marking. . . " giving an example of "the plane's landing(?)", where an upward tone at the end would indicate that the speaker is asking a question, and a static tone would indicate a statement of fact. S1 drew attention to the fact that "word stress in English can lead to two different words depending on where the stress is." An example of stress making a lexical distinction would be in between the words "permit" (noun) vs. "permit" (verb).

Aspects of vocal prosody in text is communicated via various methods such all-caps for shouting [28]. Such methods were created by a community of speakers that innovated within the constraints of the medium so that it closer resembled aspects of speech. If WhatsHap had its own long-lasting community of speakers, it would be reasonable to expect that they would discover similar innovations to convey aspects of speech beyond the phoneme.

[*Category 1C: types of communication for which the system is best suited*] — Participants S1, S2, H2, and H3 noted that they felt that haptic messaging was best suited for information-centered communication where there is a shared context between users. S2 and S1 and H3 expressed that between the two activities, the *reasoning gap* (deciding a time and activity) vs. *information gap* (coordinating a food order between a chef and supplier), they felt that the latter task was easier. S2: "this works best when you're just trying to get information from the other person. If you're trying to like decide on where you should go, or what you should do I feel like it's harder." S1 and S2 both concurred that it was difficult to use the system to gauge mood and preference. S2 further stressed that the system would work best if "you already have a real conversation happening in some other context where you're preferably in a better mode of communication, and this is more like a confirmation check." H3 felt that shorter messages were easier to interpret, so it was best if conversants who received messages haptically should send requests, and receive short confirmations due to message length.

Offering some discussion to this finding, we believe that the discrete, time-constant nature of the phoneme-based encoding system may contribute to this, as it obfuscates prosodic distinctions that have both semantic and pragmatic consequences. The use of WhatsHap with text input similarly suffers loss of information since text-based gestures such as emoji are not rendered.

*2) **Theme 2: influences on cognitive load**:* When conversing with the system, the time between replies is typically quite long compared to spoken or textual conversation, with turns taking an average of 2.84 min ($SD = 1.90\,\mathrm{min}$). This is indicative of the task having a high cognitive load (CL). It should be noted that our trained participants had only 250 min to practice. With increased training we may expect lower CL demands, as exemplified with new vs. experienced readers. These are explored in this theme's categories.

[*Category 2A: training*] — Many haptic listeners emphasized the importance of practice contributing to their stress and performance during sessions. Even S1 and S2, who did not receive messages haptically, noted that they felt like they improved both in skill and confidence after sessions. S1 noted that the later sessions went faster, reporting that there was "less waiting and thinking about 'oh was that too long of a sentence?'." H1 said they wished that they specifically trained on recognizing frequently occurring words: " 'and', 'are', 'do', 'does'. Any more practice on those things, specific words that are used a lot in conversation." This is of interest because despite their minimization, not all functional words were eliminated from conversation, and still occurred sufficiently frequently to warrant this sentiment.

[*Category 2B: mental state and performance*] Both haptic listeners and speakers stressed the importance of confidence and mental well-being in performing the task adequately. H1 had poor performance on their first session, attributing this to the fact that they were stressed and had to leave at a specific time. With increased practice, these stressful feelings were assuaged. Haptic listeners also highlighted the importance of memory when answering questions. Phonemes were delivered with 1 s inter-stimulus-intervals, with 3 s pauses between words. This lead to lengthy transmission times. H3 said "it is harder to recollect longer sentences because by the time you get to the end of the sentence you forget it."

[*Category 2D: encoding system adjustments for reduced CL*] — H3 voiced concern over high transmission times and felt that the encoding system could be adjusted for reduced mental strain. They felt they would have preferred placing pauses only between syllable and word boundaries, effectively using a syllable-based approach to deliver messages instead of a phoneme-based approach.

[*Category 2C: linguistic and cultural background*] — H3, a haptic-listener who speaks Hindi said that this system may suit some languages more than others because of cultural practice around language instruction and orthography. He mentioned that Hindi orthography was taught to him in terms of its articulatory phonetics, with glyphs having a close correspondence to speech sounds, unlike English. He felt that his cultural background made learning the system easier. S2, a fluent English speaker, mentioned how English is not typically taught in a phoneme-aware way, which may make phoneme learning more difficult for native English speakers. There are also many varieties of English that may influence a speaker's expectations. H3 noted confusion between British and American pronunciation of certain words like "classmate".

## V. Conclusion and Future Work

In this work, we contribute towards commercially ready wearable speech replacement devices by demonstrating the feasibility of haptically mediated conversation aimed at accomplishing real-world tasks using a lightweight and fully mobile apparatus. We demonstrated that participants were able to successfully complete conversational tasks most of the time (87.5%). However, interlocutors must align in context and linguistic structure during the conversation for effective communication. Communicating haptically through WhatsHap has a large influence on the participants' perceived cognitive load.

Future work for WhatsHap could facilitate haptic conversation by integrating many of the topics discussed above: a hybrid approach that combines time-varying continuous aspects of speech with the discrete nature of phonemic perception may strike a balance between intelligibility and self expression. Additionally, the timing between speech sounds is a major factor in their perception, so if stimulus delivery does not follow fixed intervals, it may result in higher intelligibility. Applying WhatsHap to different languages may yield interesting results in terms of accuracy scores and warrant different encoding techniques. The system's focus on phonemic awareness may assist in second language learning, especially in guiding a new language learner's perception of phonemes that may have otherwise been imperceivable.

## References

[1] K. Fellah and M. Guiatni, "Tactile display design for flight envelope protection and situational awareness," *IEEE transactions on haptics*, vol. 12, no. 1, pp. 87–98, 2018.

[2] P. Alirezaee, A. Weill-Duflos, J. J. Schlesinger, and J. R. Cooperstock, "Exploring the effectiveness of haptic alarm displays for critical care environments," in *2020 IEEE Haptics Symposium (HAPTICS)*, 2020, pp. 948–954.

[3] E. Pescara, V. Diener, and M. Beigl, "Vibraid: Comparing temporal and spatial tactile cues in control room environments," in *Proceedings of the 12th ACM International Conference on PErvasive Technologies Related to Assistive Environments*, ser. PETRA '19. New York, NY, USA: Association for Computing Machinery, 2019, p. 138–145. [Online]. Available: https://doi.org/10.1145/3316782.3321531

[4] F. Sorgini, R. Caliò, M. C. Carrozza, and C. M. Oddo, "Haptic-assistive technologies for audition and vision sensory disabilities," *Disability and Rehabilitation: Assistive Technology*, vol. 13, no. 4, pp. 394–421, 2018.

[5] P. Brooks and B. J. Frost, "Evaluation of a tactile vocoder for word recognition," *Journal of the Acoustical Society of America*, vol. 74, no. 1, pp. 34–39, 1983.

[6] S. Engelmann and R. Rosov, "Tactual hearing experiment with deaf and hearing subjects," *Exceptional Children*, vol. 41, no. 4, pp. 243–253, 1975.

[7] J. M. Weisenberger and M. E. Percy, "The transmission of phoneme-level information by multichannel tactile speech perception aids." *Ear and hearing*, vol. 16, no. 4, pp. 392–406, 1995.

[8] G. Luzhnica, E. Veas, and V. Pammer, "Skin reading: Encoding text in a 6-channel haptic display," in *Proc. Intl. Symposium on Wearable Computers*. ACM, 2016, pp. 148–155.

[9] G. Luzhnica and E. Veas, "Optimising encoding for vibrotactile skin reading," in *Proc. Human Factors in Computing Systems (CHI)*. ACM, 2019, pp. 1–14.

[10] M. A. Plaisier, D. S. Vermeer, and A. M. Kappers, "Learning the vibrotactile morse code alphabet," *ACM Transactions on Applied Perception (TAP)*, vol. 17, no. 3, pp. 1–10, 2020.

[11] C. M. Reed, H. Z. Tan, Z. D. Perez, E. C. Wilson, F. M. Severgnini, J. Jung, J. S. Martinez, Y. Jiao, A. Israr, F. Lau *et al.*, "A phonemic-based tactile display for speech communication," *IEEE Transactions on Haptics*, vol. 12, no. 1, pp. 2–17, 2018.

[12] N. Dunkelberger, J. Sullivan, J. Bradley, N. P. Walling, I. Manickam, G. Dasarathy, A. Israr *et al.*, "Conveying language through haptics: a multi-sensory approach," in *Proc. Intl. Symposium on Wearable Computers*. ACM, 2018, pp. 25–32.

[13] R. Turcott, J. Chen, P. Castillo, B. Knott, W. Setiawan, F. Briggs, K. Klumb, F. Abnousi, P. Chakka, F. Lau, and A. Israr, "Efficient evaluation of coding strategies for transcutaneous language communication," in *Intl. Conf. on Human Haptic Sensing and Touch Enabled Computer Applications*. Springer, 2018, pp. 600–611.

[14] M. F. de Vargas, A. Weill-Duflos, and J. R. Cooperstock, "Haptic speech communication using stimuli evocative of phoneme production," in *2019 IEEE World Haptics Conference (WHC)*. IEEE, 2019, pp. 610–615.

[15] S. Zhao, A. Israr, F. Lau, and F. Abnousi, "Coding tactile symbols for phonemic communication," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 2018, p. 392.

[16] H. Z. Tan, C. M. Reed, Y. Jiao, Z. D. Perez, E. C. Wilson, J. Jung, J. S. Martinez, and F. M. Severgnini, "Acquisition of 500 english words through a tactile phonemic sleeve (taps)," *IEEE Transactions on Haptics*, 2020.

[17] S. P. Eberhardt, L. E. Bernstein, M. E. Demorest, and M. H. Goldstein Jr, "Speechreading sentences with single-channel vibrotactile presentation of voice fundamental frequency," *The Journal of the Acoustical Society of America*, vol. 88, no. 3, pp. 1274–1285, 1990.

[18] L. E. Bernstein, M. E. Demorest, D. C. Coulter, and M. P. O'Connell, "Lipreading sentences with vibrotactile vocoders: Performance of normal-hearing and hearing-impaired subjects," *The Journal of the Acoustical Society of America*, vol. 90, no. 6, pp. 2971–2984, 1991.

[19] M. F. de Vargas, D. Marino, A. Weill, and J. R. Cooperstock, "Speaking haptically: from phonemes to phrases with a mobile haptic communication system," *IEEE Transactions on Haptics*, 2021.

[20] H.-Y. Yao and V. Hayward, "Design and analysis of a recoil-type vibrotactile transducer," *The Journal of the Acoustical Society of America*, vol. 128, no. 2, pp. 619–627, 2010.

[21] M. A. Mines, B. F. Hanson, and J. E. Shoup, "Frequency of occurrence of phonemes in conversational english," *Language and speech*, vol. 21, no. 3, pp. 221–241, 1978.

[22] P. Boersma and D. Weenink, "Praat: Doing phonetics by computer (version 6.0. 37)," *Retrieved 14.03. 2018 from http://www.praat.org/*, 2018.

[23] R. Ellis *et al.*, *Task-based language learning and teaching*. Oxford University Press, 2003.

[24] N. S. Prabhu, *Second language pedagogy*. Oxford University Press, 1987, vol. 20.

[25] K. C. Hall, B. Allen, M. Fry, S. Mackie, and M. McAuliffe, "Phonological corpustools, version 1.2.[computer program]," *Available from PCT GitHub page*, 2016.

[26] B. Hayes, *Introductory phonology*. John Wiley & Sons, 2011, vol. 32.

[27] C. E. Shannon, "A mathematical theory of communication," *Bell system technical journal*, vol. 27, no. 3, pp. 379–423, 1948.

[28] G. McCulloch, *Because Internet: Understanding the new rules of language*. Riverhead Books, 2019.