








Phonemic-based Tactile Supplement to Lipreading: Initial Findings on Consonant Identification

Charlotte M. Reed¹ , Dimitri Kanevsky² , Joseph G. Desloge³ ,
Juan S. Martinez⁴ , and Hong Z. Tan⁴ 

¹ Research Lab of Electronics, Massachusetts Institute of Technology,
Cambridge, MA 02139, USA

cmreed@mit.edu

² Google Research, Mountain View, CA 94043, USA

dkanevsky@google.com

³ San Francisco, CA, USA

⁴ Haptic Interface Research Lab, Purdue University,

West Lafayette, IN 47907, USA

{mart1304,hongtan}@purdue.edu

Abstract. We present the initial results of a 7-channel tactile aid worn on the forearm, which was designed as a supplement to lipreading. Consonant confusion patterns through lipreading alone were first assessed in one non-native English speaker with profound deafness. A phonemic-based tactile coding scheme was then developed to disambiguate sets of two or more speech sounds that appear similar on the lips (i.e., visemes). After a period of training with the tactile aid in combination with lipreading, scores for identification of a set of 22 consonants were substantially higher for the aided lipreading condition (74%-correct) than through lipreading alone (44%-correct). This performance compares favorably to that reported in the literature for spectral-based tactile aids. These initial results demonstrate the potential of a phonemic-based approach to a tactile supplement of lipreading.

Keywords: Tactile aid · Phoneme-based · Lipreading · Supplement · Deafness

1 Introduction

Based on statistics provided by the World Health Organization, there are roughly 12 to 25 million persons worldwide with profound deafness. Roughly 1 million Americans are estimated to be “functionally deaf,” that is, unable to hear normal conversation even with a hearing aid [8]. Severe-to-profound hearing impairment

Research supported by an award to MIT from Google LLC for work by C.M. Reed on “Wearable Tactile Displays.” It is also partially supported by NSF Grants IIS 1954886 (to MIT) and 1954842 (to Purdue).

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

C. Saitis et al. (Eds.): HAID 2022, LNCS 13417, pp. 25–34, 2022.

https://doi.org/10.1007/978-3-031-15019-7_3

has serious ramifications throughout the life cycle, with negative effects observed in terms of level of income, activity, education, and employment [15]. While some profoundly deaf individuals use sign language for communication, other deaf children and adults rely mainly on oral methods of communication. For those individuals who do not have sufficient residual hearing to derive benefit from traditional hearing aids, other prosthetic devices are available to provide spoken language to children and adults with profound deafness, including cochlear implants and tactual aids.

Although performance with current wearable tactual aids is often inferior to that achieved with cochlear implants [10], not all deaf persons who rely on spoken language are able to be implanted or to achieve benefits from implantation. Speech-to-text applications, now readily available for use on smartphones, also provide Deaf and Hard-of-Hearing persons access to spoken language. However, a need still exists for face-to-face communication without the use of captioning applications. There may be times when this type of technology is not available or inconvenient to use, and additionally, some members of this population may prefer to watch the face of the talker for non-verbal cues (e.g., facial expressions that convey a range of emotions) that are lost in the captioning process. With the proposed tactual aid to lipreading, the user has access to a more complete and richer version of the spoken message. Thus, it is important to continue to develop tactual aids for individuals with profound hearing impairment for stand-alone use or as a supplement to hearing aids, cochlear implants, or captioning. A tactile aid to lipreading would be of significant benefit to the quality of life and well-being of millions of people with severe to profound hearing loss.

The TActile Phonemic Sleeve (TAPS) was recently developed for speech transmission through touch alone. Training and testing data from more than 100 individuals show that English speakers can learn to use TAPS within hours of training, and the best users acquire English words at a rate of 1 word per minute with a vocabulary size up to 500 words (essentially open vocabulary) [13, 14, 16]. The present research is aimed at a pared-down version of TAPS that supplements lipreading based on a phonemic coding scheme customized for the viseme groups of an individual deaf user of the device. In this paper we describe the design of the tactile aid to lipreading, and report on experiments conducted with one deaf participant to measure consonant recognition through lipreading alone and in combination with the tactile aid.

2 Methods

The experiments reported here were conducted under a protocol approved by the Internal Review Board (IRB) of MIT, through which the participant provided informed consent.

2.1 Consonant Confusion Patterns with Lipreading only

The first step in this project was to determine the consonant confusion patterns through lipreading alone of the deaf individual for whom the tactile aid was

being designed. This participant (the second author of this paper) is an adult male who became profoundly deaf in early childhood. He is a non-native speaker of English whose first language is Russian. He does not use a hearing aid or cochlear implant, but has had previous experience with other tactile aids to lipreading (see [1]).

In the current study, the participant’s ability to identify consonants in the initial position of consonant-vowel (CV) syllables through lipreading alone was examined using materials from the STeVi speech corpus of English.¹ The test syllables consisted of audio-visual recordings of 22 consonants (/p t k b d g f θ s ʃ v ð z ʒ ʧ ʤ m n r w l j/) with the vowels /i a u/ produced by each of four talkers (2 M and 2 F) in the carrier phrase “You will mark [CV] please.” Identification tests were conducted on each talker separately using a one-interval, 22-alternative forced-choice procedure. The experiment was controlled by a Matlab program (based on the AFC software developed by [3]) running on a desktop computer. The stimuli were presented over a computer monitor and the participant’s task was to use a computer mouse to select a response from an orthographic display of the 22 stimuli that appeared after the stimulus had been presented. Each run consisted of 66 trials using random presentation without replacement of the syllables produced by a given talker. Trial-by-trial, correct-answer feedback was provided on the first run for each talker. The subsequent five runs, which employed a different repetition of the syllables, were conducted without feedback. Across the four talkers, 1320 test trials were obtained and used to construct a stimulus-response confusion matrix for data analysis summarized below.

Overall percent-correct score across talkers was 31.94% with overall information transmission of 2.031 bits. A hierarchical clustering analysis (performed in Matlab using a city-block distance measure and average linkage) yielded a dendritic tree with the following seven consonant clusters: /p b m/, /f v/, /θ ð/, /s z t d/, /ʃ ʒ ʧ ʤ/, /k g n j l/ and /r w/. These results are similar to those obtained in lipreading studies with native English speakers (see [9]), and were used to guide the development of the phonemic-based tactile aid to lipreading described below in Sect. 2.2.

2.2 Tactile Coding of Consonants to Supplement Lipreading

Our strategy in the design of a tactile code was to create tactile signals that would permit a lipreader to distinguish among the phonemes within each of the seven viseme groups identified above. The phonemes within each viseme group are assigned different tactile codes. By watching the face of the talker and attending to the tactile cue, the lipreader will be provided with sufficient information to distinguish among sounds that are highly confused through lipreading alone.

Description of Tactile Device. The tactile codes are presented through a 7-channel tactile device applied to the forearm and thenar eminence (see photo in Fig. 1). This device is a simplified version of the 24-channel phonemic-based

¹ (<https://www.sens.com/products/stevi-speech-test-video-corpus/>).



Fig. 1. Layout of factors for the 7-channel tactile aid to lipreading.

tactile aid described by [14], using similar hardware, software, and principles for the selection of tactile codes. The actuators are wide-band tactors (Tectonic Elements, Model TEAX13C02-8/RH, Part #297–214, sourced from Parts Express International, Inc.) measuring 30 mm in diameter and 2 mm in thickness. Seven channels of a 24-channel USB audio device (MOTU, model 24Ao, Cambridge, MA) were used to deliver audio waveforms to each of the tactors through custom-built audio amplifiers. Matlab programs running on a desktop computer were used to generate waveforms to drive each tactor independently and synchronously with an associated video stimulus.

The user placed the left forearm on the vibrators palm side down such that the volar surface makes contact with the tactors labeled 2, 4, and 6, and the thenar eminence of the palm with tactor 7. The piece of material containing the tactors labeled 1, 3, and 5 was placed on top of the forearm and secured with strips of cloth that wrap around the arm and held in place with Velcro. Stimulus levels used in the speech experiments were established by first obtaining measurements of tactile thresholds at one tactor (the “reference” tactor labelled “3” in Fig. 1) for the frequencies used in the coding (60, 150, 250 Hz). A loudness-matching procedure was then used to adjust the level of each tactor such that its perceived strength was equal to that of the reference tactor for a 250-Hz signal presented at a level roughly 20 dB above threshold. These level adjustments in dB relative to maximum output for each tactor were applied to the other two stimulating frequencies (60 Hz and 150 Hz), based on data reported by [19] demonstrating similar loudness-growth contours across frequency.

Mapping of Phonemes to Tactile Display The device assumes real-time phoneme recognition at its front end. However, in order to remove the con-

found of phoneme-recognition accuracy with tactile code performance, the current study obtained perfect phoneme recognition using the non-real time Montreal Forced Aligner [7]. For consonants, each phoneme was assigned a tactile code whose duration was determined by the length of time it was recognized by the alignment scheme.

Table 1. Tactile codes for 22 phonemes, arranged by viseme group.

Phoneme	Tactor	Carrier (Hz)	30-Hz AM
/b/	5	250	No
/m/	5	250	Yes
/p/			
/v/	5	250	No
/f/			
/ð/	5	250	No
/θ/			
/d/	5	250	No
/t/			
/z/	1	250	Yes
/s/	1	250	No
/ʒ/	1	250	Yes
/ʃ/	1	250	No
/ʒ/	5	250	No
/ʒ/			
/r/	6	250	No
/w/	1	250	No
/l/	1	60	No
/j/	2	250	No
/n/	5	250	Yes
/g/	2	60	No
/k/			

The tactile codes are shown in Table 1 and are described in terms of the tactor that was used for stimulation (as numbered in Fig. 1), the carrier frequency in Hz, and the presence or absence of a 30-Hz amplitude modulation (AM) applied to the carrier frequency. A different tactile code is assigned to each phoneme within a viseme group. For example, for visemes /b m p/, /b/ is coded with a 250-Hz vibration at tactor 5, /m/ with a 250-Hz carrier with amplitude modulation at a rate 30Hz at tactor 5, and /p/ is coded by the absence of a tactile signal. Among the full set of 22 consonants, codes were presented at four different locations (tactors labelled 1, 2, 5, and 6 in Fig. 1), used two different carrier

frequencies (60 and 250 Hz), and were presented either with or without a 30-Hz amplitude modulation applied to the carrier frequency. The absence of a tactile signal was also used for coding, as shown for /p f t tʃ k/ in Table 1. Note that multiple phonemes can share the same tactile code (such as /m/ and /n/), as long as they belong to different viseme groups and thus are visibly different on the lips.

Table 2. Training results for each of the 11 stimulus sets.

Set	No. of phonemes	Stimulus set	Training with correct-answer feedback %-correct score			
			Run 1	Run 2	Run 3	Run 4
1	3	/p b m/	88.9	83.3	100	
2	4	/f v θ ð/	100			
3	7	Sets 1 + 2	90.5	100		
4	2	/r w/	100			
5	9	Sets 3 + 4	94.4			
6	4	/s z t d/	66.7	79.2	70.0	87.5
7	13	Sets 5 + 6	87.2			
8	4	/ʃ ʒ tʃ ʒ/	91.7			
9	17	Sets 7 + 8	84.1			
10	5	/l j n g k/	73.3	76.7	60.3	76.0
11	22	Sets 9 + 10	78.8			

The tactile stimulus codes were presented at a level of 30 dB SL relative to the threshold measured 250 Hz on tactor 3, and then applying the level adjustments for the remaining tactors for equal perceived strength, as described in Sect. 2.2 above.

In addition to the phoneme-specific codes, an envelope signal derived from an octave band of the acoustic speech signal with center frequency 500 Hz was used to drive a 150-Hz carrier at tactor 7 at a level of 30 dB SL.

2.3 Training and Testing with Tactile Aid as a Supplement to Lipreading

The participant received training in the use of the tactile aid in combination with lipreading, using the CV speech materials as described in Sect. 2.1 for the lipreading alone study. The participant was instructed to watch the lips of the talker and at the same time attend to the tactile code that was presented during a given stimulus presentation. The 22 consonants were grouped into sets for training (based primarily on viseme groups), and then gradually combined into larger groups over the course of the training. The eleven stimulus sets that were

employed in the training are shown in the first three columns of Table 2. The sets are numbered from 1 through 11, as shown in the first column. The second and third columns provide the number of phonemes in the set and the identity of the phonemes, respectively. All stimuli introduced up to a given point in the training process are represented in sets 3, 5, 7, 9, and 11.

Training included the following steps for each stimulus set: The participant was shown a schematic diagram of the stimulus codes and felt them in isolation; the stimuli were presented in CV syllables in a fixed order; and closed-set identification tests were conducted with trial-by-trial feedback. When performance exceeded 80% correct or had leveled off, testing was conducted without feedback. The participant was given the option of repeating stimulus presentations before responding to focus on the integration of the tactile cue with lipreading. The identification tests were conducted as described in Sect. 2.1 adjusted to test only the stimuli in the set with the number of trials presented on the training and test runs updated to reflect the set. For stimulus sets with 2 to 13 phonemes (sets 1 through 8 and set 10), each run consisted of 6 randomized presentations of each consonant. For the larger stimulus sets (set 9 with 17 stimuli and set 11 with 22 stimuli), each run consisted of 3 randomized presentations of each consonant.

3 Consonant Recognition Results with Tactile Supplement

Results of training on the condition of lipreading with the tactile supplement are summarized in Table 2.

The %-correct score is shown after each run of trials with correct-answer feedback for the eleven sets of stimuli used in training. The number of training runs was dependent on performance as described above and thus varied across stimulus sets. Perfect performance of 100%-correct was achieved with training on the first 4 stimulus sets. For Sets 5 through 9, scores in the range of 84 to 94%-correct were achieved at the end of training. The most difficult viseme group was Set 10 where performance reached a plateau of 76% correct after 4 training runs. Performance for the full set of consonants in Set 11 reached 79% correct after one training run.

Following training with feedback, scores on each stimulus set were then obtained for tests conducted without the use of feedback. These scores were compared to scores obtained on each stimulus set for the condition of lipreading alone. In Fig. 2, %-correct scores are shown comparing lipreading alone with aided lipreading for each of the eleven stimulus sets, as defined in Table 2. The %-correct scores for lipreading alone is shown by the height of the blue bars, and the scores for the condition of lipreading combined with the tactile aid are shown by the height of the red bars. Performance with the tactile aid exceeded that with lipreading alone by at least 30% points for each of the sets. On Set 11, which included the full set of 22 consonants, an improvement of 31% points was obtained for the combined condition of lipreading plus the tactile aid (74.2%-correct) compared to lipreading alone (43.9%-correct).

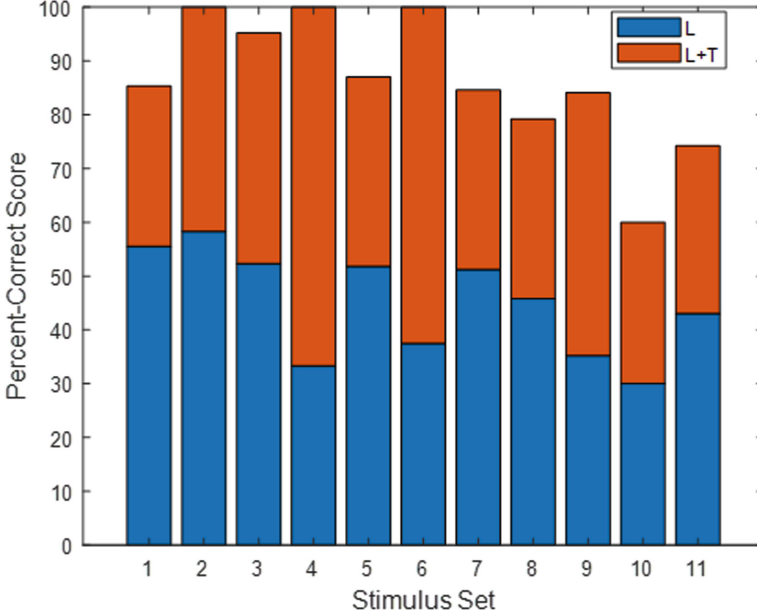


Fig. 2. Percent-correct scores for each of the 11 stimulus sets (as defined in Table 2) showing performance on lipreading alone (blue bars) and lipreading plus tactile aid (red bars). (Color figure online)

4 Discussion and Concluding Remarks

These preliminary studies with a phoneme-based tactile supplement to lipreading have shown promising results for improving the recognition of consonants for the deaf participant who received training with the aid. The results shown here for the benefits of the tactile aid compare favorably to those shown in previous evaluations of tactile aids (e.g., [4, 20]). The results reported here, however, are limited to consonant recognition with one deaf participant. Further evaluations are necessary to include recognition of vowels as well as consonants, and to include a larger sample of persons with hearing impairment in the evaluations. Nonetheless, the results obtained in this preliminary evaluation are promising for continued work on the development of a wearable system.

Much of the previous work on the development of tactile aids has focused on a spectral-based approach to the processing and display of the acoustic speech signal. These devices employ spectral decomposition of the acoustic speech signal for presentation of different spectral bands to different places on the skin (e.g., see reviews in [5, 11, 12]). In contrast to spectral-based processing, recent work on tactile aids [e.g., [2, 6, 13, 14, 16–18, 21]] has focused on a phoneme-based approach to encoding speech signals. This approach assumes the use of automatic speech recognition (ASR) at the front-end of the device for real-time recognition of phonemes. One important contrast between the spectral-based and phoneme-

based approaches lies in the manner in which the inherent variability in speech tokens within and across talker is handled. In spectral-based displays, the burden of interpreting this variability falls in the tactile domain on the user of the device. In the phoneme-based approach, however, this task is accomplished by the ASR component, allowing the association of a specific tactile code to each phoneme. The use of an invariant code for each phoneme should be conducive to the process of integrating tactile signals with lipreading.

Further research on phoneme-based tactile supplements to lipreading is underway to extend the results of the preliminary study reported here. Ongoing work includes the design and evaluation of phoneme-based tactile codes for aided lipreading of vowels, as well as continued training and testing of deaf participants. Work is also planned for the development of a wearable version of this device. Such a tactile aid to lipreading would encompass real-time processing of the acoustic speech signal for extraction of features (such as voicing and nasality) that would then be encoded on a tactile array.

Acknowledgments. The authors wish to thank Dick Lyon, Pascal Getreuer, Artem Dementyev, and Malcolm Slaney of Google Research for their valuable discussions and feedback on this research.

References

1. Cholewiak, R.W., Sherrick, C.E.: Tracking skill of a deaf person with long-term tactile aid experience: a case study. *J. Rehabil. Res. Dev.* **23**(2), 20–26 (1986)
2. Dunkelberger, N., et al.: A multisensory approach to present phonemes as language through a wearable haptic device. *IEEE Trans. Haptics* **14**(1), 188–199 (2020)
3. Ewert, S.D.: AFC - A modular framework for running psychoacoustic experiments and computational perception models. In: *Proceedings of the International Conference on Acoustics AIA-DAGA*, pp. 1326–1329. Merano, Italy (2013)
4. Galvin, K.L., Ginis, J., Cowan, R.S., Blamey, P.J., Clark, G.M.: A comparison of a new prototype tickle talkerTM with the tactaid 7. *Aust. N. Z. J. Audiol.* **23**(1), 18–36 (2001)
5. Kappers, A.M., Plaisier, M.A.: Hands-free devices for displaying speech and language in the tactile modality-methods and approaches. *IEEE Trans. Haptics* **14**(3), 465–478 (2021). <https://doi.org/10.1109/TOH.2021.3051737>
6. Martinez, J.S., Tan, H.Z., Reed, C.M.: Improving tactile codes for increased speech communication rates in a phonemic-based tactile display. *IEEE Trans. Haptics* **14**(1), 200–211 (2020)
7. McAuliffe, M., Socolof, M., Mihuc, S., Wagner, M., Sonderegger, M.: Montreal forced aligner. *Comput. Prog. Version 0.9.0* (2017). Accessed 17 Jan 2017. <http://montrealcorpusools.github.io/Montreal-Forced-Aligner/>
8. Mitchell, R.E.: How many deaf people are there in the United States? Estimates from the survey of income and program participation. *J. Deaf Stud. Deaf Educ.* **11**(1), 112–119 (2005). <https://doi.org/10.1093/deafed/enj004>
9. Owens, E., Blazek, B.: Visemes observed by hearing-impaired and normal-hearing adult viewers. *J. Speech Lang. Hear. Res.* **28**(3), 381–393 (1985)
10. Reed, C.M., Delhorne, L.A.: Current results of a field study of adult users of tactile aids. *Semin. Hear.* **16**(4), 305–315 (1995)

11. Reed, C.M., Durlach, N.I., Braida, L.D.: Research on tactile communication of speech: a review. *ASHA Monogr.* **20**, 1–23 (1982)
12. Reed, C.M., Durlach, N.I., Delhorne, L.A., Rabinowitz, W.M., Grant, K.W.: Research on tactual communication of speech: Ideas, issues, and findings. In: McGarr, N.S. (eds.) *The Volta Review (Monograph entitled “Research on the Use of Sensory Aids for Hearing-Impaired People”)* (1989)
13. Reed, C.M., Tan, H.Z., Jiao, Y., Perez, Z.D., Wilson, E.C.: Identification of words and phrases through a phonemic-based haptic display: effects of inter-phoneme and inter-word interval durations. *ACM Trans. Appl. Percept.* **18**(3) (2021). <https://doi.org/10.1145/3458725>
14. Reed, C.M., et al.: A phonemic-based tactile display for speech communication. *IEEE Trans. Haptics* **12**(1), 2–17 (2019). <https://doi.org/10.1109/TOH.2018.2861010>
15. Ries, P.W.: Prevalence and characteristics of persons with hearing trouble, United States, 1990–91. National Center for Health Statistics. Series 10: Data from the National Health Survey No. 188 (1994)
16. Tan, H.Z., et al.: Acquisition of 500 English words through a TActile phonemic sleeve (TAPS). *IEEE Trans. Haptics* **13**(4), 745–760 (2020). <https://doi.org/10.1109/TOH.2020.2973135>
17. Turcott, R., et al.: Efficient evaluation of coding strategies for transcutaneous language communication. In: Prattichizzo, D., Shinoda, H., Tan, H.Z., Ruffaldi, E., Frisoli, A. (eds.) *EuroHaptics 2018*. LNCS, vol. 10894, pp. 600–611. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-93399-3_51
18. de Vargas, M.F., Weill-Duflos, A., Cooperstock, J.R.: Haptic speech communication using stimuli evocative of phoneme production. In: 2019 IEEE World Haptics Conference (WHC 2019), pp. 610–615. IEEE (2019)
19. Verrillo, R.T., Fraioli, A.J., Smith, R.L.: Sensation magnitude of vibrotactile stimuli. *Percept. Psychophysics* **6**(6), 366–372 (1969). <https://doi.org/10.3758/BF03212793>
20. Weisenberger, J.M., Percy, M.E.: The transmission of phoneme-level information by multichannel tactile speech perception aids. *Ear Hear.* **16**(4), 392–406 (1995). <https://doi.org/10.1097/00003446-199508000-00006>
21. Zhao, S., Israr, A., Lau, F., Abnoui, F.: Coding tactile symbols for phonemic communication. In: *Proceedings of the 2018 ACM CHI Conference on Human Factors in Computing Systems*, pp. 1–13 (2018)