35th Annual International Conference of the IEEE EMBS
Osaka, Japan, 3 - 7 July, 2013

# The role of augmentative visual training in auditory human-machine-interface performance*

Gabrielle L. Hands, Eric Larson, and Cara E. Stepp

*Abstract*— **The purpose of this study was to evaluate the effect of augmentative visual feedback training on performance using auditory feedback alone for human-machine interface (HMI) control. Sixteen healthy participants used bilateral facial surface electromyography to achieve two-dimensional control to reach vowel targets. Eight participants trained with combined visual and auditory feedback, while eight participants trained with real-time auditory feedback only. Each subject participated in four sessions over three days; three sessions with their designated feedback modality (auditory only or auditory with supplementary visual) and a fourth session on the third day using novel vowel targets to test generalization of auditory-motor learning. Analyses of variance performed on the percentage of total targets reached demonstrated a main effect of group and the interaction of group and session. Individuals provided with augmentative visual feedback during training outperformed individuals using auditory feedback alone in initial training sessions. However, training with augmentative visual feedback had no effect on individuals' training and generalization performance using auditory feedback alone after three days of training.**

## I. INTRODUCTION

Human-machine interfaces (HMI) serve as augmentative communication pathways that allow users to control external devices. Many currently use electroencephalography (EEG) or surface electromyography (sEMG) to transmit signals produced by the brain or muscles into control of an interface [1-3]. HMIs are primarily useful for patients with little remaining motor function, such as those suffering from severe spinal cord injury, amytrophic lateral sclerosis (ALS), or locked-in syndrome (LIS).

Many of the current HMI designs call for constant visual contact with the user, often requiring the user to control their eye movement and shift their gaze during the task. In order to successfully control HMIs, it is imperative that the user receives feedback on their performance. The majority of the HMIs in practice have implemented visual feedback [4], as it offers high performance rates and is easy for new users to learn. However, a constant visual connection is demanding for all users and infeasible for patients with ALS or LIS without intact vision. In addition, more mistakes were observed during control of a visual HMI when paired with the presentation of distracting visual stimuli among control subjects [5]. These findings suggest that other feedback modalities must be explored in order to make HMIs more user-friendly and practical as communication support.

To address the proposed issues with visually guided HMIs, several studies have used interfaces that are controlled using only listening paradigms through evoked brain responses measured by EEG (e.g., P300) [6-8]. Although these paradigms show relatively high performance, systems that rely on evoked responses are inherently slow. Alternatively, several studies have examined the feasibility of HMIs that present continuous auditory feedback to the user [3, 9-12]. These paradigms have led to mixed results, but the auditory-only groups in these studies performed consistently worse than groups that used visual feedback.

It is not clear, however, that *combined* auditory-visual feedback during continuous HMI control improves performance. As in other combinations of sensory modalities (notably visual-haptic), performance may be dependent upon the context of the task [13] and exact formulation of each modality. For instance, a study of the effects of feedback modality on the control of an HMI that utilized subjects' ability to self-regulate SCPs found the smallest learning effect in the group that received combined auditory-visual feedback [14]. Additionally, Guenther et al. [9] studied the ability of a single individual with LIS to control vowel production using an implanted brain electrode. Participants were asked to move in the auditory "vowel space" from a central vowel location to one of three peripheral vowel locations (/i/ in "beat", /a/ in "pot", or /u/ in "boot"). During 10 of the training sessions, no visual feedback was provided to the subject, whereas in another 15 sessions augmentative two-dimensional visual feedback was provided. The authors anecdotally noted no difference in performance between presentation of auditory-only and auditory-visual feedback during operation of a speech synthesizer. However, given the invasive nature of this paradigm, no formal study could be accomplished in a larger group of users to establish the generalizability of this finding. Furthermore, only performance in reaching *trained* targets was compared.

Although the benefits of audio-visual feedback during HMI control are questionable, the role of augmentative visual feedback during *training* on auditory feedback has not been studied. Here we examine the role of training with and without visual augmentative feedback on both performance and generalization of auditory HMI control. The task consisted of trying to achieve American English vowel targets (auditory feedback) through two-dimensional control of bilateral facial sEMG. The participants were divided into

---

*Research supported by the Boston University Grants for Undergraduate Teaching and Scholarship Program.

G.L. Hands is with the Department of Speech, Language, and Hearing Sciences, Boston University, Boston, MA 02215 USA (e-mail: ghands@bu.edu).

E. Larson is with the Department of Speech and Hearing Sciences, University of Washington, Seattle, WA 98105 USA (e-mail: larsoner@uw.edu).

C.E. Stepp is with the Departments of Speech, Language, and Hearing Sciences and Biomedical Engineering, Boston University, Boston, MA 02215 USA (phone: 617-353-7487; fax: 617-353-5074; e-mail: cstepp@bu.edu).

978-1-4577-0216-7/13/$26.00 ©2013 IEEE

an auditory only (AO) group that received only auditory feedback throughout training and testing and an auditory-visual (AV) group that received auditory and augmentative visual feedback during training.

## II. METHODS

### A. Participants

Sixteen healthy young adults participated in the experiment. All subjects were native English speakers and reported no history of speech, language, hearing or neurological disorders. Participants were pseudorandomly assigned to one of two experimental groups: auditory-visual (AV) or auditory only (AO). The average age of the eight individuals (5 males) in the AV group was 20.3 years (STD=2.6) and the average age of the eight individuals (4 males) in the AO group was 19.4 years (STD=1.8). All participants completed written consent in compliance with the Boston University Institutional Review Board.

### B. Experimental Set-up

Bilateral activation of the orbicularis oris muscles was measured using a surface electromyography (sEMG) system (Delsys 2-channel Bagnoli system) and an external sound card (M-Audio Fast Track PRO) in order to control user movement in two dimensions. The Delsys system band-pass filtered sEMG signals prior to acquisition with corner frequencies of 20 Hz and 450 Hz. Double differential electrodes were placed on the left and right orbicularis oris muscles (Fig. 1), which were chosen to provide high signal-to-noise ratios for two-dimensional control of the interface. The sEMG signals were sampled at 44.1 kHz using an M-Audio external soundcard.

### C. Software Set-up and Experimental Paradigm

Custom software written in C++ translated the power of each sEMG channel into formant production and provided auditory and visual feedback. F1 values were limited between 300 and 1200 Hz and were controlled by activation of the right orbicularis oris muscle. F2 values were limited between 300 and 4000 Hz and were controlled by activation of the left orbicularis oris muscle. The sEMG power ranges were calibrated to each participant prior to the start of the experiment. Specifically, signals were recorded while participants were instructed to alternate between rest and maximum voluntary contraction (MVC). The participant's maximum and minimum power for each channel were then used to map activity onto formant locations. In the software, a two-dimensional viewing space in which the x-axis corresponded to F1 values and the y-axis corresponded to F2 values was presented to the user. The F1 axis and the F2 axis were linearly mapped to 10% - 85% for right and left sEMG channels, respectively.

Each group (AO and AV) participated in four sessions comprised of 120 trials per session over three days, each session lasting approximately 40 minutes. The first and second (training) sessions were completed on days 1 and 2, respectively. The third (training) and fourth (generalization) sessions were completed consecutively on day 3. Auditory cues and feedback were produced using a Klatt synthesizer as implemented in the STK toolkit [15]. Visual targets and feedback varied based on the session and the group (see below). Auditory feedback was provided by a loudspeaker placed in front of the participant.

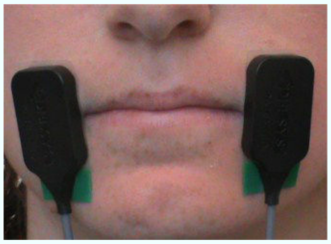

Figure 1. Electrode placements over the right and left orbicularis oris.

### D. AV Group

In the auditory-visual (AV) group, participants were instructed to contract their left and right orbicularis oris muscles in order to reach a target vowel sound and corresponding ellipse in the two-dimensional viewing window (see Fig. 2). The three training targets used for sessions 1, 2 and 3 were fixed as ellipses in the F1-F2 plane associated with the American English vowels /I/, /u/ and /a/. These targets corresponded to the cue words "bit", "boot" and "pot", respectively. Each vowel target was presented 40 times per session. During session 4 (generalization), novel target sounds were presented to the participant. These targets were also fixed as ellipses in the F1-F2 plane and were associated with the American English vowels /i/, /æ/ and /o/ with the cue words "beat", "bat" and "boat", respectively. In all sessions, AV subjects received real-time auditory feedback as to their location in the F1-F2 plane in addition to the presentation of auditory and visual target presentation. The level of visual feedback presented to the user during training depended on the session. In all sessions the participants' goal was to modulate their sEMG activation to produce the target vowel in the F1-F2 plane, represented by an ellipse that covered a range of F1 and F2 values for that

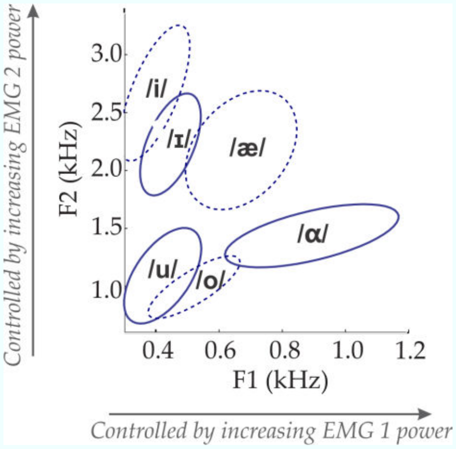

Figure 2. Methods for 1) training (solid lines) and 2) test/generalization (dashed lines) using established American English vowel categories. Target locations are shown as ellipses in the formant (F1-F2) plane. Solid ellipses designate the vowel targets trained in the first three sessions, while the dashed ellipses designate novel (untrained) vowel targets participants had to generalize to in the final (fourth) session. Participants moved in 2D by manipulating power in two EMG locations. Increasing EMG 1 power led to increases in formant 1 whereas increasing EMG 2 power led to increases in formant 2.

particular vowel sound. Participants had 15 seconds to reach the target vowel ellipse before the session timed out.

In session 1, in addition to online auditory feedback, AV participants received real-time visual feedback in the form of a cursor that moved across the visual representation of the F1-F2 plane based on their sEMG activation. Participants received an auditory and visual target before each trial: an auditory cue of the vowel sound, the presentation of a fixed sample word that contained the sound, and presentation of an ellipse that corresponded to the F1 and F2 values associated with the target vowel. Participants also received target visual feedback in the form of an ellipse located in the two-dimensional space and presented the word cue in the center of the two-dimensional space during the trial. When the participant correctly localized the vowel target, the ellipse turned darker in color, signaling task success. During the first training session, participants were instructed to attend to the auditory cues in order to help them locate the target once they had adjusted to manipulating the sEMG. For session 2, the real-time visual feedback cursor was removed and only real-time auditory feedback was used. In addition to the auditory target, the visual target ellipse was still presented and still turned darker in color when the subject achieved the target. Participants were again encouraged to pay attention to the real-time auditory feedback in order to locate the target. For session 3, all real-time visual and target visual feedback was removed. Target ellipses were never visually presented. The participants were still presented with the word cue in the middle of the two-dimensional space throughout the session. In this session only, the auditory cue was presented at the beginning of the trial to designate the target and the participant had to use only real-time auditory feedback to locate the target. No feedback was provided regarding whether or not the participant had reached the target, so participants were instructed to hold their sEMG activation when they thought they had achieved the correct target vowel sound. Session 4 was identical to session 3, except that participants were presented novel vowel targets (/i/, /o/, and /æ/) for auditory-motor generalization instead of the targets that they had trained with. As in session 3, participants could only use continuous auditory feedback and the presentation of the word cue in the middle of the screen to locate the target; use of novel targets required them to perform auditory-motor generalization. In all sessions, the trial ended when the target was reached or after 16 seconds of attempts.

### *E. AO Group*

Training and testing of the AO group was similar to that of the AV group. The first three sessions (training sessions 1 - 3) of the AO group were identical to training session 3 of the AV group: participants received only real-time auditory feedback and a centered visual word cue to locate their vowel target. Participants were instructed to use the auditory cue heard at the beginning of the trial as a target and modulate their sEMG activation in order to replicate this auditory cue. Real-time auditory feedback was presented so that participants could hear their movement in the two-dimensional space and locate the target. Neither visual target ellipses nor moving visual cursors were ever shown or discussed with AO participants. Session 4 of the AO group was identical to session 4 of the AV group: the participants were presented with novel vowel targets (/i/, /o/, and /æ/) for auditory-motor generalization.

### *F. Data Analysis*

Participants' success was measured based on their ability to achieve the vowel target locations in each trial. Performance was calculated as the percent of total trials in each session in which the subjects achieved the target locations using custom MATLAB software (Mathworks, Natick, MA). Statistical analysis was performed using Minitab Statistical Software (Minitab Inc, State College, PA). A two-factor mixed models analysis of variance (ANOVA) was performed to determine the effect of feedback type (between-subjects, auditory only vs. auditory-visual), session (within-subjects, 1-4) and the interaction of feedback type × session. *Post hoc* Two-tailed student t-tests were performed across session and within group (paired) as well as within session and across group (unpaired) using a Bonferroni Correction (28 total comparisons, $p_{adj} = 0.0017$).

## III. RESULTS

Figure 3 shows the mean performance as a function of session for the two feedback groups. In session 1, participants using full auditory and visual feedback (AV group) were able to achieve an average performance of 98.3% (STD = 2.8) whereas participants using auditory feedback alone (AO group) were only able to achieve 47.5% of the targets (STD = 21.0). In session 2, participants using full auditory and partial visual feedback (AV) were able to reach 87.9% performance (STD = 2.1) and those using auditory feedback alone (AO) were able to reach 50.8% (STD = 20.5). In the final training session (session 3), both groups used auditory feedback alone and performed the same task with the same feedback (see Methods). Here the AV group reached an average of 80.2% performance (STD = 15.9) and the AO an average of 56.1% performance (STD = 24.0). Finally, in session 4, both groups were asked to reach new, untrained targets using only auditory feedback. The AV group averaged 77.7% performance (STD = 17.9) whereas the AO group averaged performance of 66.0% (STD = 15.6).

The results of the ANOVA indicated a significant effect of group ($p < 0.001$) and the interaction of group × session ($p < 0.001$), but not of session ($p > 0.05$). The effect of group showed a large effect size ($\eta_p^2 = 0.62$); individuals in the AV group showed higher performance than those in the AO group. The interaction of group × session showed a moderate-to-large effect size ($\eta_p^2 = 0.34$). *Post hoc* t-tests within session and across groups found significant differences ($p < 0.0017$) between the AO and AV group at sessions 1 and 2, but not at sessions 3 and 4. No differences were seen as a function of session number for either group.

## IV. DISCUSSION

Individuals provided with augmentative visual feedback outperformed individuals using auditory feedback alone in initial sessions. However, after training, use of augmentative visual feedback training in early sessions had no effect on individuals' ability to use the auditory HMI to attain training or test targets. This result matches well with the previous

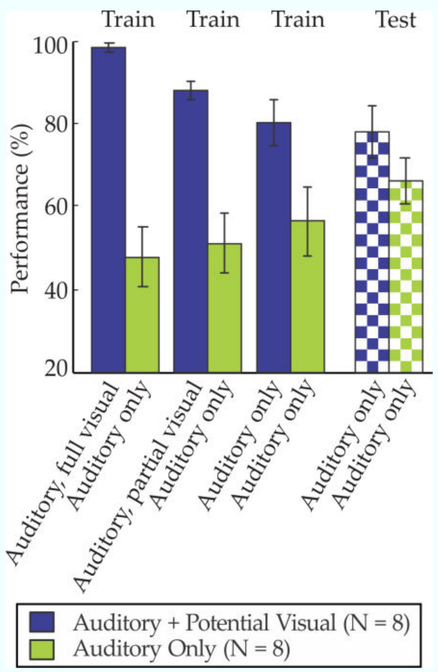

Figure 3. Results of 1) training (solid fill) and 2) test/generalization (checkered fill) The performance (% targets reached) is shown during three training sessions (solid fill) with varying levels of augmentative visual feedback (dark blue) or always using only auditory feedback (light green). A final generalization/test session is shown in checkered fill using novel targets and auditory only feedback. *Post hoc* Tukey's tests across the interactions between group and session found significant differences ($p < 0.05$) between the AO and AV group at sessions 1 and 2, but not at sessions 3 and 4. Error bars indicate +/- the standard error of the mean.

report of Guenther et al., who did not note a difference in the ability of an individual with LIS to control an auditory HMI when given augmentative visual feedback [9].

Although no significant differences were seen within group as a function of session, our data show a trend for individuals in the AV group to reduce performance as the visual feedback is taken away and for individuals in the AO group to increase performance with training. The previous work of Nijboer compared individuals trained over three training sessions of EEG HMI control using either auditory feedback or visual feedback [10]. Although individuals in the auditory group showed overall lower performance, only one of the eight participants did not show improvement with training. However, six of the eight participants in the visual group had lower performance during their last block of training relative to their first block. These results indicate that with visual feedback, participants have strategies immediately available to regulate EEG whereas auditory feedback seems to retard learning. This difference could potentially be related to the effects of subject mood and motivation, in which good initial performance using the visual feedback led to increased mood and confidence, in later sessions digressing into an increased fear of incompetence and thus decreased performance.

Use of augmentative visual feedback may improve initial auditory HMI performance, likely due to users' reliance on vision, which alone has been shown to provide increased HMI control [10, 12]. However, here we have shown that use of augmentative visual feedback during training does not increase individuals' performance using auditory feedback alone – using both training and test targets. Future development of auditory HMIs need not incorporate visual feedback training to achieve auditory-motor learning.

## ACKNOWLEDGMENT

The authors thank Elias Thorp (currently of Northwestern University) and Alan Pacheco of Boston University.

## REFERENCES

[1] S. P. Kelly, E. C. Lalor, C. Finucane, G. McDarby, and R. B. Reilly, "Visual spatial attention control in an independent brain-computer interface," *IEEE Trans Biomed Eng,* vol. 52, pp. 1588-96, Sep 2005.

[2] E. W. Sellers and E. Donchin, "A P300-based brain-computer interface: initial tests by ALS patients," *Clin Neurophysiol,* vol. 117, pp. 538-48, Mar 2006.

[3] E. D. Larson, H. Terry, and C. E. Stepp, "Audio-visual feedback for electromyographic control of vowel synthesis," presented at the Engineering in Medicine and Biology Society (EMBC) Annual International Conference of the IEEE, San Diego, CA, 2012.

[4] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain-computer interfaces for communication and control," *Clin Neurophysiol,* vol. 113, pp. 767-91, Jun 2002.

[5] F. Cincotti, L. Kauhanen, F. Aloise, T. Palomaki, N. Caporusso, P. Jylanki, D. Mattia, F. Babiloni, G. Vanacker, M. Nuttin, M. G. Marciani, and R. M. J. Del, "Vibrotactile feedback for brain-computer interface operation," *Comput Intell Neurosci,* p. 48937, 2007.

[6] H. Higashi, T. M. Rutkowski, Y. Washizawa, A. Cichocki, and T. Tanaka, "EEG auditory steady state responses classification for the novel BCI," *Conf Proc IEEE Eng Med Biol Soc,* vol. 2011, pp. 4576-9, 2011.

[7] M. A. Lopez-Gordo, E. Fernandez, S. Romero, F. Pelayo, and A. Prieto, "An auditory brain-computer interface evoked by natural speech," *J Neural Eng,* vol. 9, p. 036013, Jun 2012.

[8] M. Schreuder, B. Blankertz, and M. Tangermann, "A new auditory multi-class brain-computer interface paradigm: spatial hearing as an informative cue," *PLoS ONE,* vol. 5, p. e9813, 2010.

[9] F. H. Guenther, J. S. Brumberg, E. J. Wright, A. Nieto-Castanon, J. A. Tourville, M. Panko, R. Law, S. A. Siebert, J. L. Bartels, D. S. Andreasen, P. Ehirim, H. Mao, and P. R. Kennedy, "A wireless brain-machine interface for real-time speech synthesis," *PLoS ONE,* vol. 4, p. e8218, 2009.

[10] F. Nijboer, A. Furdea, I. Gunst, J. Mellinger, D. J. McFarland, N. Birbaumer, and A. Kubler, "An auditory brain-computer interface (BCI)," *J Neurosci Methods,* vol. 167, pp. 43-50, Jan 15 2008.

[11] F. Oscari, R. Secoli, F. Avanzini, G. Rosati, and D. J. Reinkensmeyer, "Substituting auditory for visual feedback to adapt to altered dynamic and kinematic environments during reaching," *Exp Brain Res,* vol. 221, pp. 33-41, Aug 2012.

[12] M. Pham, T. Hinterberger, N. Neumann, A. Kubler, N. Hofmayer, A. Grether, B. Wilhelm, J. J. Vatine, and N. Birbaumer, "An auditory brain-computer interface based on the self-regulation of slow cortical potentials," *Neurorehabil Neural Repair,* vol. 19, pp. 206-18, Sep 2005.

[13] C. E. Stepp, B. T. Dellon, and Y. Matsuoka, "Contextual effects on robotic experiments of sensory feedback for object manipulation," presented at the Biomedical Robotics and Biomechatronics (BioRob), 2010 3rd IEEE RAS and EMBS International Conference on, Tokyo, Japan, 2010.

[14] T. Hinterberger, N. Neumann, M. Pham, A. Kubler, A. Grether, N. Hofmayer, B. Wilhelm, H. Flor, and N. Birbaumer, "A multimodal brain-based feedback and communication system," *Exp Brain Res,* vol. 154, pp. 521-6, Feb 2004.

[15] G. S. a. P. Cook, "RtMIDI, RtAudio, and a Synthesis (STK) Update," *Proceedings of the International Computer Music Conference, Barcelona, Spain,* 2005.