Human Movement Science 35 (2014) 145–155

Contents lists available at ScienceDirect

Human Movement Science

journal homepage: www.elsevier.com/locate/humov

Effects of augmentative visual training on audio-motor mapping

Gabrielle L. Hands [a], Eric Larson [b], Cara E. Stepp [a,c,*]

[a] Department of Speech, Language, and Hearing Sciences, Boston University, Boston, MA, USA
[b] Institute for Learning and Brain Sciences, University of Washington, Seattle, WA, USA
[c] Department of Biomedical Engineering, Boston University, Boston, MA, USA

ARTICLE INFO

Article history:
Available online 12 February 2014

Keywords:
Auditory–motor
Visual feedback
Surface electromyography

ABSTRACT

The purpose of this study was to determine the effect of augmentative visual feedback training on auditory–motor performance. Thirty-two healthy young participants used facial surface electromyography (sEMG) to control a human–machine interface (HMI) for which the output was vowel synthesis. An auditory-only (AO) group ($n = 16$) trained with auditory feedback alone and an auditory–visual (AV) group ($n = 16$) trained with auditory feedback and progressively-removed visual feedback. Subjects participated in three training sessions and one testing session over 3 days. During the testing session they were given novel targets to test auditory–motor generalization. We hypothesized that the auditory–visual group would perform better on the novel set of targets than the group that trained with auditory feedback only. Analysis of variance on the percentage of total targets reached indicated a significant interaction between group and session: individuals in the AV group performed significantly better than those in the AO group during early training sessions (while using visual feedback), but no difference was seen between the two groups during later sessions. Results suggest that augmentative visual feedback during training does not improve auditory–motor performance.

---

* Corresponding author at: Department of Speech, Language, and Hearing Sciences, Boston University, 635 Commonwealth Avenue, Boston, MA 02215, USA. Tel.: +1 617 353 7487.

E-mail address: cstepp@bu.edu (C.E. Stepp).

http://dx.doi.org/10.1016/j.humov.2014.01.003
0167-9457/

1. 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 translate signals produced by the brain or muscles into control of an interface (Larson, Terry, Canevari, & Stepp, 2013; Sellers & Donchin, 2006; Trejo, Rosipal, & Matthews, 2006). HMIs using these neurophysiological signals are primarily useful for patients with little remaining motor function, such as those suffering from spinal cord injury, amytrophic lateral sclerosis (ALS) or locked-in syndrome (LIS).

Several different classes of HMIs currently exist in research settings: passive, in which the output of the interface is determined by the involuntary brain activity of the user, and active, in which the user deliberately controls the interface and requires feedback on their performance (Zander, Kothe, Jatzev, & Gaertner, 2010; Zander, Kothe, Welke, & Roetting, 2008). The majority of both active and passive HMIs currently in use call for constant visual monitoring, requiring the user to control their eye movement and shift their gaze during a task. In active HMIs specifically, it is imperative that the user receives feedback on their performance in order to successfully control the interface. Many of the current HMI designs have implemented visual feedback 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 is infeasible for patients with ALS or LIS who do not have intact vision. In addition, more mistakes were observed during control of a visual HMI when paired with the presentation of distracting visual stimuli among healthy participants (Cincotti et al., 2007). These findings suggest that alternative feedback modalities should be explored in order to make HMIs more user-friendly and practical as a means of communication support. To address the feasibility of removing the visual channel from HMI designs, several studies have proposed both passive and active interfaces to be controlled with the aid of auditory feedback.

Passive designs using auditory stimuli have shown feasibility, typically employing listening paradigms that result in evoked brain responses that are measured using EEG (e.g., P300; Higashi, Rutkowski, Washizawa, Cichocki, & Tanaka, 2011; Lopez-Gordo, Fernandez, Romero, Pelayo, & Prieto, 2012; Schreuder, Blankertz, & Tangermann, 2010). Higashi et al. (2011) measured auditory steady-state EEG responses while participants attended to a tone that was presented to the left or right in each trial, and were able to evoke discriminable EEG responses simply by attending to auditory stimuli on either side. Lopez-Gordo et al. (2012) used an EEG-based BCI that implemented human voice in a similar dichotic listening paradigm and found group average classification accuracy as high as 70%. Another auditory BCI that used spatial hearing as a cue suggested that healthy participants could also use auditory spatial attention to elicit P300 responses for classification (Schreuder et al., 2010). The results of these studies have demonstrated that healthy individuals can control HMIs using auditory stimuli alone. Although these paradigms show relatively high performance, systems that rely on evoked potentials are inherently slow. More research into *active* auditory-only designs is needed.

A few studies have compared users' abilities to control active HMIs using auditory versus visual feedback (Guenther et al., 2009; Larson et al., 2013; Nijboer et al., 2008; Oscari, Secoli, Avanzini, Rosati, & Reinkensmeyer, 2012; Pham et al., 2005). These paradigms have led to mixed results, but the auditory-only groups in these studies performed consistently worse than the groups that received additional visual feedback. For instance, Nijboer et al. (2008) trained healthy subjects to continuously control the amplitude of their EEG sensorimotor rhythms using either auditory or visual feedback. While the average success rate of the participants who received visual feedback was 70%, only half of the participants in the auditory-only group could reach 70% at any point in their training. In addition, the auditory-only group required longer training time on their respective task than the visual group. Pham et al. (2005) examined the ability of healthy participants to control slow cortical potentials (SCPs) when using either auditory or visual feedback. While overall performance was similar between the auditory and visual groups, the experimenters reported that responses in the auditory-only group were more variable; the auditory-only group was less able to self-regulate SCPs. These results suggest that while healthy subjects can learn to use auditory feedback to control active HMIs, they generally have more difficulty controlling HMIs when presented with auditory feedback.

One explanation for the previously lackluster results using auditory feedback may be that auditory feedback is less intuitive than visual feedback. If so, a combination of auditory–visual feedback could lead to improved performance. However, as in other combinations of sensory modalities, performance may be dependent on the context (Stepp, Dellon, & Matsuoka, 2010) and exact formulation of the task. For instance, a study of the effects of feedback modality on control of an HMI that utilized subjects' ability to self-regulate SCPs found that performance was significantly lower in the group that received only auditory feedback compared to the group that received only visual feedback, but the smallest learning effect was seen in the group that received *both* auditory and visual feedback (Hinterberger et al., 2004). Conversely, Guenther et al. (2009) studied the ability of one individual with LIS to control vowel production using an implanted brain electrode. The participant was asked to move in the auditory "vowel space" from a central vowel location to one of three peripheral vowel locations (/i/ in "beat", /ɑ/ 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 observed no difference in performance between presentation of auditory-only and auditory–visual feedback during operation of a speech synthesizer. Finally, Klobassa et al. (2009) compared the effects of auditory–visual cues and auditory cues alone on control of a P300-based HMI and found that both groups demonstrated similar performances. This may suggest that combining the two feedback modalities does not aid in the learning of the HMI task. These results, taken in concert, suggest that combined auditory and visual feedback may not improve HMI control compared to auditory feedback alone.

Although the benefits of audio–visual feedback during HMI operation are questionable, the role of augmentative visual feedback during *training* on auditory feedback has not yet been studied. If auditory feedback is not intuitive for users, using visual feedback during HMI training may make eventual auditory-only device control more straightforward for users. This potential is tempered by the findings of previous studies indicating that subjects may learn most effectively with the feedback modality that they practice with and that the visual feedback modality tends to be most heavily relied on (Coull, Tremblay, & Elliott, 2001; Khan & Franks, 2000; Proteau, Marteniuk, & Levesque, 1992). Here we examine the role of training with and without augmentative visual feedback on both performance and generalization of auditory HMI control. We seek to address the extent to which auditory feedback alone during training (compared to audio–visual training) is sufficient for healthy subjects to control an auditory-only active HMI. The present study expanded upon a previously designed auditory-based HMI (Larson et al., 2013) and compared a group that received only auditory feedback throughout training to a group receiving combined auditory and visual feedback throughout training. The task consisted of trying to produce American English vowel targets (which served as auditory feedback) using two-dimensional control via bilateral facial sEMG. Participants were assigned into either an auditory-only (AO) group that received only auditory feedback throughout training and testing, or an auditory–visual (AV) group that received auditory and visual feedback during training, but was tested using auditory feedback only. Both groups received the same vowel sounds throughout training and were tested on the same set of novel vowel sounds during the last session with only auditory feedback. We hypothesized that the group that received augmentative visual feedback during training might perform better (overall obtaining a higher percentage of vowel targets) on the novel vowel set in the last session than the group that trained using only auditory feedback.

2. Methods

2.1. Participants

Thirty-two 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 16 individuals (5 males) in the AV group was 20.1 years (STD = 2.1); the average age of the 16 individuals in the AO group (7 males) was 19.7 years (STD = 1.6). All participants completed written consent in compliance with the Boston University Institutional Review Board.

2.2. Experimental set-up

Bilateral activation of the orbicularis oris muscles was measured using a surface electromyography (sEMG) system (Delsys two-channel Bagnoli system) in combination with an external sound card (M-Audio Fast Track PRO. The Delsys Bagnoli system band-pass filtered sEMG signals prior to acquisition with corner frequencies of 20 and 450 Hz. The bilateral orbicularis oris muscles were chosen to provide high signal-to-noise ratios for control of the interface. Before placing the sEMG electrodes on the subject, the skin was prepared by swabbing with alcohol wipes and exfoliating with tape (peeling) according to previous methods (Stepp, 2012). Double differential electrodes were cleaned, prepared with double-sided adhesive interfaces, and placed over the left and right orbicularis oris muscles. A ground electrode (Dermatrode) was placed on the acromion process of each subject. Each channel was amplified by a gain of 1000 prior to acquisition. The sEMG signals were then digitized at 44.1 kHz.

2.3. Software set-up and experimental paradigm

Custom software written in C++ translated the power of each sEMG channel into auditory and visual feedback. Auditory feedback was in the form of voice synthesis produced using a Klatt synthesizer as implemented in the STK toolkit (Cook & Scavone, 1999) in which the first two vowel formants (F1 and F2) were determined by the power in the two sEMG channels. 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 range of power of each sEMG signal was calibrated for each participant prior to the start of the experiment. Specifically, sEMG signals were recorded while participants were instructed to alternate between rest and maximum voluntary contraction (MVC). The sEMG signals were windowed using a 2048-sample (46 ms) Hanning window with no overlap and then smoothed over time using a decaying exponential filter with a 1 s time constant in order to remove noise from the sEMG measurements. The participant's maximum and minimum power for each channel during the calibration task were then used to map sEMG activity onto locations in the F1–F2 plane. The F1 axis and the F2 axis were linearly mapped to 10–85% of MVC for right and left sEMG channels, respectively. Additionally, 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 members of the AV group during selected sessions.

Members of each group (AO and AV) participated over 3 days in four sessions comprised of 120 trials each, with each session lasting approximately 40 min. 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 feedback was provided by a loudspeaker placed in front of the participant. Visual targets and feedback presented to the members of the AV group varied based on the session (see below).

2.4. AV group

Here we will outline the training and testing structure for the AV group; while the experimental parameters for the AO group were similar, we will specify the differences in a later section (see Section 2.5). Participants in the AV group were instructed to coordinate contraction of their left and right orbicularis oris muscles in order to reach a target vowel sound and corresponding ellipse in the two-dimensional viewing window. The three training targets used for sessions 1, 2, and 3 were ellipses in the F1–F2 plane associated with the American English vowels /I/, /u/ and /ɑ/, corresponding 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 (see Fig. 1). 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 (see Fig. 1). In all sessions, participants received real-time auditory feedback as to their location in the

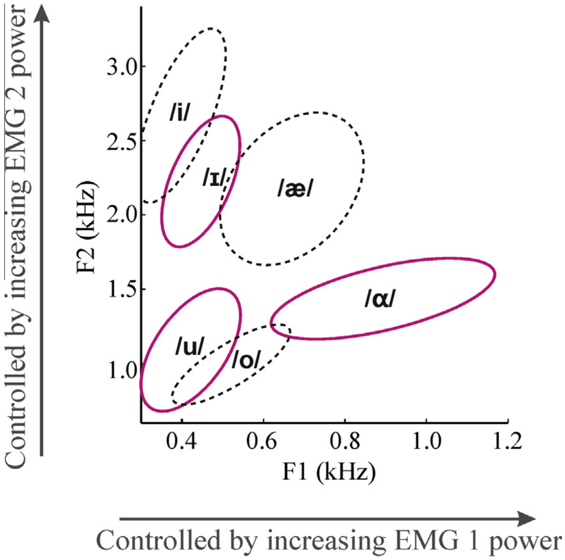

Fig. 1. Methods for training set and test/generalization set of American English vowel targets. Target locations are shown as ellipses in the F1–F2 plane. Solid lines indicate the training set of vowel sounds used during the first three training sessions and dashed lines indicate the novel vowel targets presented in the fourth (final) test session. Participants controlled movement in the F1–F2 plane by manipulating their movement in two EMG directions. Increasing activation of EMG electrode 1 caused increases in formant 1 while increasing activation of EMG 2 caused increases in formant 2.

F1–F2 plane in addition to the presentation of the auditory targets. 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 particular vowel sound. Participants had 15 s to reach the target vowel ellipse before the trial timed out.

The level of visual feedback presented to the user during training depended on the session. In session 1, participants received both an auditory and visual target before each trial: an auditory cue of the vowel sound, the presentation of a sample word that contained the sound fixed in the center of the screen, and visual presentation of an ellipse that corresponded to the F1 and F2 values associated with the target vowel. In addition to online auditory feedback (a synthetic vowel sound with continuously changing F1 and F2 values), in session 1 AV participants also 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. Finally, participants also received "knowledge of results" visual feedback: when the participant correctly localized the vowel target, the ellipse turned darker in color, signaling task success. During this 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. During session 2, the real-time visual feedback cursor was removed and only real-time auditory feedback was used. However, 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 to help them locate the target. Finally, during sessions 3 and 4, all real-time visual and target visual feedback was removed. Target ellipses were never visually presented. The participants were presented with the word cue in the middle of the two-dimensional space throughout the session, as well as an auditory cue at the beginning of the trial and only had access to 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 differed from session 3 only in that participants were presented with novel vowel targets (/i/, /o/, and /æ/) instead of the targets that they had trained with in sessions 1–3. Use of novel targets required participants to perform auditory–motor generalization.

2.5. 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 session 3 of the AV group: participants received only an auditory cue of the vowel sound, a centered visual word cue, and real-time auditory feedback 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 participants. As in the AV group, vowels /I/, /u/ and /ɑ/ were used as targets in these sessions. 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 but were not given any visual targets or feedback other than the centered visual word cue located in the middle of the screen.

2.6. Data analysis

Participants' success was measured based on their ability to achieve the vowel target locations. Vowel target locations were successfully reached by correct modulation of participants' sEMG activation to produce the target vowel in the F1–F2 plane. 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, AO vs. AV), 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), correcting for the multiple (16) comparisons using a conservative Bonferroni correction (alpha = .05/16 = .0031).

3. Results

A trial was considered successful when the participant correctly modulated their EMG activation to produce the target vowel in the F1–F2 plane. Each session consisted of 120 trials. Performance measures suggest that while subjects with auditory-only (AO) feedback could not perform as well as those with auditory–visual (AV) feedback during training, their performance during testing was not significantly different from that of the AV group (Fig. 2). The results of the ANOVA (Table 1) indicated a significant effect of session ($p = .005$), group ($p < .001$) and the interaction between session and group ($p < .001$) on performance (success rate). The effect of group showed a moderate-to-large effect size ($\eta_p^2 = .35$), with individuals in the AV group achieving higher performance overall than members of the AO group. However, the interaction between session and group also showed a moderate-to-large effect size ($\eta_p^2 = .33$). *Post hoc t*-tests within session and across groups found significant differences between the AO and AV group in sessions 1 and 2 (both $p < .0031$) (Tables 2 and 3), but not in sessions 3 and 4. Within the AV group, significant differences in performance were found between session 1 and session 2, session 1, and session 3, and session 1 and session 4. No differences were seen as a function of session within the AO group or for any other comparisons within the AV group.

In session 1, participants in the AV group performed at an average of 96.8% (STD = 4.8%) whereas participants in the AO group achieved an average of 50.8% (STD = 18.2%). In session 2, participants in the AV group reached an average of 80.5% (STD = 19.2%) and participants in the AO group achieved an average of 47.5% (STD = 21.7%). In training session 3, neither members of the auditory-only group nor the auditory–visual group received visual feedback about their performance (see Section 2). In this

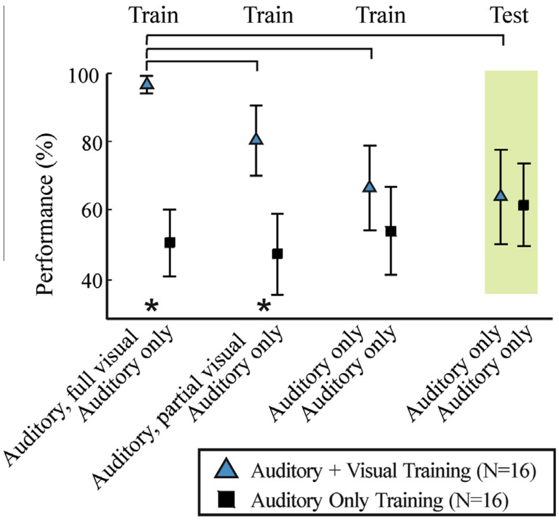

Fig. 2. Results of training and test/generalization sessions. The performance (% of total targets reached) during each training session with varying levels of augmentative visual feedback is shown in light blue fill for the auditory–visual group and black fill for the auditory-only group. *Post-hoc* tests determined significant differences between the AO and AV groups during sessions 1 and 2 (shown with asterisks), but no difference was found during sessions 3 and 4. Performance of the AV group during session 1 was significantly higher relative to sessions 2–4 (shown via brackets). Error bars indicate the standard error of the mean. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1
Results of ANOVA for group, session, and the interaction between group and session. Significant results are marked with an asterisk.

Factor	DF	η_p^2	*F*	*p*
Group	1.0	.349	16.1	*<.001
Session	2.4	.147	5.2	*.005
Group × session	2.4	.329	14.7	*<.001
Error	90.0			

Table 2
Results (*p*-values) of *post hoc t*-tests across session (within each group). Significant results (based on the Bonferroni-corrected threshold alpha = .0031) are marked with an asterisk.

	Session 1 vs. session 2	Session 1 vs. session 3	Session 1 vs. session 4	Session 2 vs. session 3	Session 2 vs. session 4	Session 3 vs. session 4
AV	*.0008	*.0001	*.0001	.0175	.0099	.4420
AO	.5628	.5488	.0540	.1886	.0314	.0801

Table 3
Results (*p*-values) of *post hoc t*-tests comparing AV and AO groups (within session).

Session 1	Session 2	Session 3	Session 4
*<.0001	*.0001	.1421	.7899

session, the AV group achieved an average of 66.7% (STD = 23.1%) and participants in the AO group performed at an average of 52.2% (STD = 23.8%). In session 4, both groups were instructed to reach novel, untrained targets using only auditory feedback. In this session, the AV group averaged 64.2% (STD = 25.8%) and the AO group averaged 61.8% (STD = 22.3%).

4. Discussion

This study examined the effects of augmentative visual feedback during training on auditory–motor HMI control. Our results indicate that although the AV group was able to control the HMI with increased performance relative to the AO group during training sessions 1 and 2 (while visual feedback was present), no differences were seen between the AV and AO groups when *only* auditory feedback was present. This was true for training targets (session 3) as well as generalization targets (session 4). These results confirm the anecdotal observations of Guenther et al. (2009), who saw no difference in the performance of an individual with LIS controlling a speech synthesizer using an implanted brain electrode when using auditory-only or auditory–visual feedback.

4.1. Augmentative visual feedback during training did not improve HMI control

Although use of visual feedback during task performance (e.g., sessions 1 and 2) allowed participants in the AV group to reach high levels of performance in HMI control, the AV group did not out-perform the AO group after visual feedback was taken away in sessions 3 and 4. We hypothesized that the AV group might perform better in the last sessions than the group that trained using only auditory feedback, with the early visual feedback giving a scaffold to the otherwise difficult task. Conversely, it is possible that members of the AV group may have developed an over-reliance on the visual channel during sessions 1 and 2 that counteracted their ability to leverage any benefits of visual feedback, with the decrement in their performance over sessions specifically related to the differences between vision-alone and auditory-alone feedback This could be related to the findings of previous studies that learning of a task is specific to the conditions provided during training (Coull et al., 2001; Khan & Franks, 2000; Proteau et al., 1992). Proteau et al. (1992) trained subjects to displace a stylus to a target despite added disruption of their movement. Subjects were divided into two groups: one performed the task in a dark room (no visual feedback) and the other performed the task in a bright room (visual feedback). They found that both groups performed significantly worse during transfer tests when they were tested on the feedback modality that they did not train with. Coull et al. (2001) also found that subjects' ability to learn across trials is dependent upon the feedback modality that they practice with. They found that a group that received visual feedback on a force production task for 10 trials and subsequently switched to auditory feedback performed significantly worse following the modality change. These results may suggest that subjects tend to develop an over-reliance on vision when presented visual and auditory feedback. Tremblay and Proteau (1998) suggest that this drop in performance is because visual feedback is the dominant source of information. Finally, Khan and Franks (2000) examined the effect of providing visual feedback during training and subsequently removing the visual feedback before the participants knew how well they were performing in an elbow rotation task. Their results suggest that not only did vision improve subjects' accuracy during training, but removal of the visual feedback negatively affected performance even after practice. Taken in concert, these results suggest that learning depends on training modality and that subjects may rely most heavily on the visual modality during learning of a task, which may have resulted in an over-dominance on the visual feedback in the AV group.

In this study, some care was taken to remove visual feedback gradually (real-time cursor and target position in session 1; target position only in session 2), the visual feedback was always *reliable*. Previous work suggests that information from the visual channel supersedes haptic information unless the visual feedback channel is sufficiently unreliable (Ernst & Banks, 2002). Similarly, although vision frequently trumps conflicting auditory perception, recent work indicates that this is consistent with a similar scheme of weighting sensory channel reliability (Battaglia, Jacobs, & Aslin, 2003; Witten & Knudsen, 2005). Future work to examine the use of augmentative visual feedback in auditory–motor performance should investigate the utility of slowly increasing the variability (error) of the visual information rather than simply removing it. It is possible that this will leverage the hypothesized variance-minimizing perceptual processing to enable a more natural transition to auditory-only feedback.

A potential alternative interpretation of the relatively poor performance in sessions 3 and 4 for the AV group is that although the augmentative visual feedback may have been useful, concurrent decreases in motivation could have masked any of these effects. Garris, Ahlers, & Driskell (2002) have proposed a model for learning in which the essential characteristic of effective learning is an appropriate cycle of judgment–behavior-feedback. Specifically, the learner should receive positive feedback, which leads to more motivated behavior, which leads to better learning. This suggests that difficult new tasks should start easy (allowing for positive feedback) and slowly become more difficult. Similarly, the seminal work of Lawrence (1952) suggests that transitioning from an 'easy' task to a more 'difficult' one improves learning. That study looked at rats' ability to discriminate between two similar shades of grey when different groups of rats were trained with either (1) the most difficult discrimination case, or (2) an ordered continuum of transition difficulty from easy to hard, or (3) an abrupt transition from the easiest discrimination to the most difficult. The study found that the group trained on difficult stimuli showed the slowest learning, and that the group that trained on the ordered difficulty performed the most accurately on the test sessions. Further, the group that transitioned abruptly from easiest to hardest showed greater accuracy than the hard transition group but less than the continuous one. These results suggest that the best way to learn a difficult task is to start training on easier discriminations and slowly transition into more difficult ones. In our study the AV group underwent a gradual transition, designed to correspond most closely with the continuum training group in the Lawrence study. However, we did not find that this paradigm resulted in higher accuracy after training when compared to the AO group, which most similarly mimicked the high-difficulty group in the Lawrence study. This could be due to other experimental factors (outlined previously) or to inherent differences in these findings between perceptual-motor and purely perceptual learning. Conversely, in our AV group, participants started with high performance, but knew that the task would become more difficult, which may have resulted in a fear of failure that decreased their motivation. This result is similar to the findings of Nijboer et al. (2008), who found that individuals using only visual feedback showed a trend for decreased performance over three sessions of training. Future studies could conduct surveys on each day inquiring about the subjects' feelings on the experiment so far and their motivation at each point.

4.2. Performance of the auditory-only group

No statistically significant differences in performance were seen as a function of session within the AO group. However, the AO group did show an overall trend of improvement with each day of training, starting at 50.8% (STD = 18.2%) in session 1 and reaching 61.8% (STD = 22.3%) in session 4. This trend was, however, small and not statistically significant, potentially due to a number of factors. One potential reason for the lack of improvement could be a ceiling in HMI control performance due to not understanding the auditory–motor task. Alternatively, this lack of improvement could have been due to non-salient auditory feedback. Learned vowel categories display the perceptual magnet effect such that stimuli within the category are more easily categorized and less discriminable by listeners [13]. However, the boundaries of these perceptual categories differ (slightly) from person to person. More salient auditory feedback may be created by modifying targets to more precisely match individuals' vowel categories. This "tuning" to individual auditory categories could improve auditory–motor control performance and should be explored in the future.

5. Strengths and limitations

This is the first study to characterize the effect of augmentative visual feedback during training on performance of auditory active HMI control. Strengths of this study are the use of a relatively large sample of subjects and a generalization (test) set of targets, neither of which are common in HMI research. Because training on a novel task was of interest, intra-subject design could not be used, which limits the power of the study. Furthermore, the results are specific to the type of auditory feedback used here, which is not commonly employed in HMI designs. Future work should determine whether these findings generalize to alternative auditory–motor mappings. Here we provided some insight about auditory–motor performance, with and without augmentative visual aids; future work to examine the alternative case (visuo-motor performance, with and without augmentative auditory aids) is necessary to determine the bases of our findings. Finally, the control modality (sEMG) is likely to provide a higher signal-to-noise ratio than some other modalities conventionally employed in HMIs (e.g., EEG), so overall performance would likely change based on the control modality employed.

6. Conclusions

We have shown that augmentative visual feedback does not have an appreciable effect on healthy participants' abilities to control an auditory-only HMI, with no evidence supporting inclusion of augmentative visual feedback in future HMI experiments. No effect of session was seen in the auditory-only training group, suggesting that auditory–motor generalization can be achieved without extensive training. These findings have implications for future auditory–motor HMI paradigms, which may provide more user-friendly external communication devices for individuals with motor impairment that don't require constant employment of the visual channel.

Acknowledgements

The authors thank Elias Thorp and Alan Pacheco for assisting with participant recordings and analysis. Portions of this work were funded by NIH NIDCD Grants T32DC000018 (EDL), 1F32DC012456 (EDL) and 1R03DC012651 (CES), and a Boston University Grant for undergraduate teaching and scholarship (CES).

References

Battaglia, P. W., Jacobs, R. A., & Aslin, R. N. (2003). Bayesian integration of visual and auditory signals for spatial localization. *Journal of the Optical Society of America A – Optics, Image Science, and Vision, 20*, 1391–1397.

Cincotti, F., Kauhanen, L., Aloise, F., Palomaki, T., Caporusso, N., Jylanki, P., et al (2007). Vibrotactile feedback for brain–computer interface operation. *Computational Intelligence and Neuroscience*, 48937.

Cook, P. R., & Scavone, G. P. (1999). The Synthesis ToolKit (STK). *Proceedings of the International Computer Music Conference*, 164–166.

Coull, J., Tremblay, L., & Elliott, D. (2001). Examining the specificity of practice hypothesis: Is learning modality specific? *Research Quarterly for Exercise and Sport, 72*, 345–354.

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature, 415*, 429–433.

Garris, R., Ahlers, R., & Driskell, J. E. (2002). Games, motivation, and learning: A research and practice model. *Simulation & Gaming, 33*(4), 441–467.

Guenther, F. H., Brumberg, J. S., Wright, E. J., Nieto-Castanon, A., Tourville, J. A., Panko, M., et al (2009). A wireless brain–machine interface for real-time speech synthesis. *PLoS One, 4*, e8218.

Higashi, H., Rutkowski, T. M., Washizawa, Y., Cichocki, A., & Tanaka, T. (2011). EEG auditory steady state responses classification for the novel BCI. *Conference Proceedings: IEEE Engineering in Medicine and Biology Society, 2011*, 4576–4579.

Hinterberger, T., Neumann, N., Pham, M., Kubler, A., Grether, A., Hofmayer, N., et al (2004). A multimodal brain-based feedback and communication system. *Experimental Brain Research, 154*, 521–526.

Khan, M. A., & Franks, I. M. (2000). The effect of practice on component submovements is dependent on the availability of visual feedback. *Journal of Motor Behavior, 32*, 227–240.

Klobassa, D. S., Vaughan, T. M., Brunner, P., Schwartz, N. E., Wolpaw, J. R., Neuper, C., et al (2009). Toward a high-throughput auditory P300-based brain–computer interface. *Clinical Neurophysiology, 120*, 1252–1261.

Larson, E., Terry, H. P., Canevari, M. M., & Stepp, C. E. (2013). Categorical vowel perception enhances the effectiveness and generalization of auditory feedback in human–machine-interfaces. *PLoS One, 8*, e59860.

Lawrence, D. H. (1952). The transfer of a discrimination along a continuum. *Journal of Comparative and Physiological Psychology, 45*, 511–516.

Lopez-Gordo, M. A., Fernandez, E., Romero, S., Pelayo, F., & Prieto, A. (2012). An auditory brain–computer interface evoked by natural speech. *Journal of Neural Engineering, 9*, 036013.

Nijboer, F., Furdea, A., Gunst, I., Mellinger, J., McFarland, D. J., Birbaumer, N., et al (2008). An auditory brain–computer interface (BCI). *Journal of Neuroscience Methods, 167*, 43–50.

Oscari, F., Secoli, R., Avanzini, F., Rosati, G., & Reinkensmeyer, D. J. (2012). Substituting auditory for visual feedback to adapt to altered dynamic and kinematic environments during reaching. *Experimental Brain Research, 221*, 33–41.

Pham, M., Hinterberger, T., Neumann, N., Kubler, A., Hofmayer, N., Grether, A., et al (2005). An auditory brain–computer interface based on the self-regulation of slow cortical potentials. *Neurorehabilitation and Neural Repair, 19*, 206–218.

Proteau, L., Marteniuk, R. G., & Levesque, L. (1992). A sensorimotor basis for motor learning: Evidence indicating specificity of practice. *Quarterly Journal of Experimental Psychology. A, Human Experimental Psychology, 44*, 557–575.

Schreuder, M., Blankertz, B., & Tangermann, M. (2010). A new auditory multi-class brain–computer interface paradigm: Spatial hearing as an informative cue. *PLoS One, 5*, e9813.

Sellers, E. W., & Donchin, E. (2006). A P300-based brain–computer interface: Initial tests by ALS patients. *Clinical Neurophysiology, 117*, 538–548.

Stepp, C. E. (2012). Surface electromyography for speech and swallowing systems: Measurement, analysis, and interpretation. *Journal of Speech Language and Hearing Research, 55*, 1232–1246.

Stepp, C. E., Dellon, B. T., & Matsuoka, Y. (2010). Contextual effects on robotic experiments of sensory feedback for object manipulation. In *Biomedical Robotics and Biomechatronics (BioRob), 2010 3rd IEEE RAS and EMBS international conference on* (pp. 58–63). Tokyo, Japan.

Trejo, L. J., Rosipal, R., & Matthews, B. (2006). Brain–computer interfaces for 1-D and 2-D cursor control: Designs using volitional control of the EEG spectrum or steady-state visual evoked potentials. *IEEE Transactions on Neural Systems and Rehabilitation Engineering, 14*, 225–229.

Tremblay, L., & Proteau, L. (1998). Specificity of practice: The case of powerlifting. *Research Quarterly for Exercise and Sport, 69*, 284–289.

Witten, I. B., & Knudsen, E. I. (2005). Why seeing is believing: Merging auditory and visual worlds. *Neuron, 48*, 489–496.

Zander, T. O., Kothe, C., Jatzev, S., & Gaertner, M. (2010). Enhancing human–computer interaction with input from active and passive brain–computer interfaces. *Human Computer Interaction Series*, 181–199.

Zander, T. O., Kothe, C., Welke, S., & Roetting, M. (2008). Enhancing human-machine systems with secondary input from passive brain-computer interfaces. In *Proceedings of the 4th international BCI workshop and training course* (pp. 144–149). Graz: Graz University of Technology Publishing House.