Proceedings of Meetings on Acoustics

Volume 21, 2014 http://acousticalsociety.org/

167th Meeting of the Acoustical Society of America
Providence, Rhode Island
5 - 9 May 2014

Session 1aPP: Psychological and Physiological Acoustics

1aPP17. Discrete and continuous auditory feedback based on pitch and spatial lateralization for human-machine-interface control

Sylvain Favrot, Carolyn M. Michener and Cara E. Stepp*

***Corresponding author's address: Speech, Language, and Hearing Sciences, Boston University, 635 Commonwealth Ave, Boston, Massachusetts 02215, cstepp@bu.edu**

The purpose of this study was to investigate auditory-motor learning via discrete and continuous auditory feedback using pitch and spatial lateralization. Sixteen subjects used facial surface electromyography (sEMG) to control a human-machine-interface (HMI). The output of the HMI was a lateralized harmonic tone. The fundamental frequency and lateralization (left-right ear) changed with the sum and the difference of the bilateral muscle signals. For 8 participants these changes were continuous, whereas the other 8 participants received discrete feedback, in which the frequency of the tone was one of 9 possible semitones (from midi note #76 to #87) and the lateralization was either left, center or right. Participants trained over three days. A mixed-models analysis of variance showed a significant effect of learning over sessions and a trend for increased performance for the group utilizing discrete feedback. Overall, information transfer rates using this purely auditory feedback averaged 38.5 bit/min by day 3, which is similar to results from similar systems utilizing visual feedback. These results show that with minimal training, auditory feedback can provide usable HMI control.

Published by the Acoustical Society of America through the American Institute of Physics

Received 26 Mar 2014; published 31 Mar 2014

1. Introduction

Human-machine interfaces (HMIs) can be used to control augmentative and alternative communication (AAC) devices, which allow the user to communicate non-verbally. HMIs can be used to assist patients with limited communication or motor skills such as individuals with high spinal cord injuries or amyotrophic lateral sclerosis (ALS). Many HMIs employ electro-encephalography (EEG) or surface electromyography (sEMG), translating these signals from the brain or muscles into commands for control of the interface (E. D. Larson, 2012; Donchin, Mar 2006; S. P. Kelly, Sep 2005). In achieving this type of real-time closed-loop control, feedback from the HMI to the user is essential. This feedback is typically provided using visual or auditory cues. Studies comparing the use of simple auditory feedback strategies such as amplitude and frequency changes have shown that auditory feedback results in substantially worse HMI control when compared with visual feedback of a cursor in two dimensions (Pham *et al.*, 2005; Nijboer *et al.*, 2008). However, there are some substantial disadvantages of employing visual feedback. It requires constant visual attention with HMI control performance drastically decreasing in the presence of distracting visual stimuli (Cincotti *et al.*, 2007). Furthermore, visual feedback is particularly challenging for individuals with visual impairment. Further exploration of auditory cues is necessary to develop effective interfaces.

Recent studies to expand the scope of auditory stimuli have shown that spatial cues may work as effective auditory cues (M. Schreuder, 2010). The goal of this study was to determine the effectiveness of using spatial cues and tones as an auditory feedback method for HMI control and to specifically determine the effects of providing continuous feedback in which pitch and spatial lateralization are mapped continuously to users' motor output versus discrete feedback in which the continuous motor commands were mapped to discrete tones and locations.

Using facial sEMG, participants controlled an HMI to reproduce a cue with a specific frequency and spatial cue in either a discrete or continuous paradigm, with the discrete paradigm providing a cue of one of nine possible semitones and a spatial lateralization of left, right, or center. Participants were trained with five target sounds over three sessions on three days.

2. Methods

2.1 Participants

Sixteen healthy young adults participated in the experiment. All were native English speakers and none of the participants reported history of speech, language or hearing disorders. Participants were assigned to one of two experimental groups: discrete or continuous. The average age of the eight discrete lateralization participants (3 male) was 20.5 (SD = 2.5) and the average age of the eight continuous lateralization participants (3 male) was 21.6 (SD = 3.0). All participants completed written consent in compliance with the Boston University Institutional Review Board.

2.2 Experimental Set-up

Two double differential sEMG electrodes (Delsys DE3.1) were placed bilaterally over the left and right side of the orbicularis oris muscle such that the electrode bars were perpendicular to the muscle fibers (see **Fig. 1**). The ground electrode was placed on the spinous process of C7. The skin was prepared by exfoliation with alcohol to reduce the skin-electrode impedance (Hermens, 1999). The sEMG signals were pre-amplified (1000×) and filtered using a Delsys Bagnoli system (Delsys, Boston MA) with low and high cutoffs of 20 Hz and 450 Hz, respectively. sEMG signals were digitalized and sampled at 44100 Hz using a sound card (M-Audio Fast Track PRO). sEMG signals were windowed using a 2048-sample Hanning window. The power estimates of the left and right sEMG channels within each window were mapped to the two control dimensions using the maximum voluntary contractions (MVCs) of the left and right side of the orbicularis oris muscle. The MVCs were calculated as the maximum power from the left and right sEMG channel over three maximal contractions. The two controlled dimensions were mapped to 10% – 85% of the MVC for each of the right and left sEMG channels. The sEMG signals were smoothed over time using a decaying exponential filter with a 1 second time constant.

Figure 1. sEMG electrodes were placed over the right and left side of the orbicularis oris muscles.

2.3 Auditory Mapping

The test software was written in C++ and translated the power of the sEMG signals into auditory feedback. Auditory feedback was provided to the participant using headphones at a comfortable listening level. The first dimension corresponded to the *sum* of the left and right smoothed sEMG signals and controlled the frequency of tone proportionally to the semitone scale, from MIDI note #76 to #87 (corresponding to a frequency from 659 Hz to 1245 Hz). It consisted of a simple tone generated by the Synthesis ToolKit (Scavone and Cook, 2005). This control was dependent on whether the participant was part of the continuous group (on the semitone scale) or the discrete group (ten notes). This simple harmonic tone was lateralized proportionally in the second dimension to the *difference* between the smoothed left and right sEMG channels. The lateralization combined (i) an overall level difference between the left and right headphone channel and (ii) a time difference between the left and right headphone channel. The maximum level difference was 20dB and the time difference was implemented using the “delayA” function from the Synthesis ToolKit (reaching a maximum of 680 ms). This headphone channel difference produced an inside-the-head percept of the tone that appeared on the side of the highest activated muscle. The lateralization was mapped to a -90° to 90° scale corresponding to a left and right percept respectively. Participants in the continuous group experienced lateralization on a continuous scale, whereas participants in the discrete group were limited to three categories: full left (-90°), center (0°) and full right (90°).

2.4 Experimental Design

Participants completed three sessions over three consecutive days. At the beginning of each session, the participant ran a calibration procedure to determine his or her MVC using custom MATLAB software (Mathworks, Natick, MA). Each session lasted between 30-45 minutes. Participants were trained during sessions to manipulate their sEMG to achieve a set of five quadrilateral targets (see **Fig. 2**).

Each session consisted of 25 trials for each of five targets. At the beginning of each trial, the target tone (with corresponding pitch and lateralization cues) was played for two seconds. Following a two second pre-trial gap, the participant was given 15 seconds to find the target solely using real-time audio feedback. The trial was completed when the participant reached the target or the 15 second time limit was reached. At the end of trial, participants were presented with visual feedback related to their speed and accuracy in reaching the target.

2.5 Data Analysis

The information transfer rates (ITRs) for each participant were calculated for each session. Average accuracy and selection rate (the inverse of the average time to reach targets) were calculated offline using MATLAB and used to define the ITR. The information rate in bits / selection is determined as: $\ln(N) + A \ln(A) + (1-A) \ln(1-A / N-1)$, in which N represents the number of possible outputs (targets) and A the accuracy of selection. The ITR (bits / unit time) is the product of the information rate (bits / selection) and the selection rate (selections / unit time). Selection rate is defined as the rate of selections the HMI user can make in a unit of time, which here is determined by the speed of the movement to the target. Higher ITRs are thus achieved by greater number of targets (N), higher accuracy (A), and faster selection rates. ITRs were calculated in bit/min.

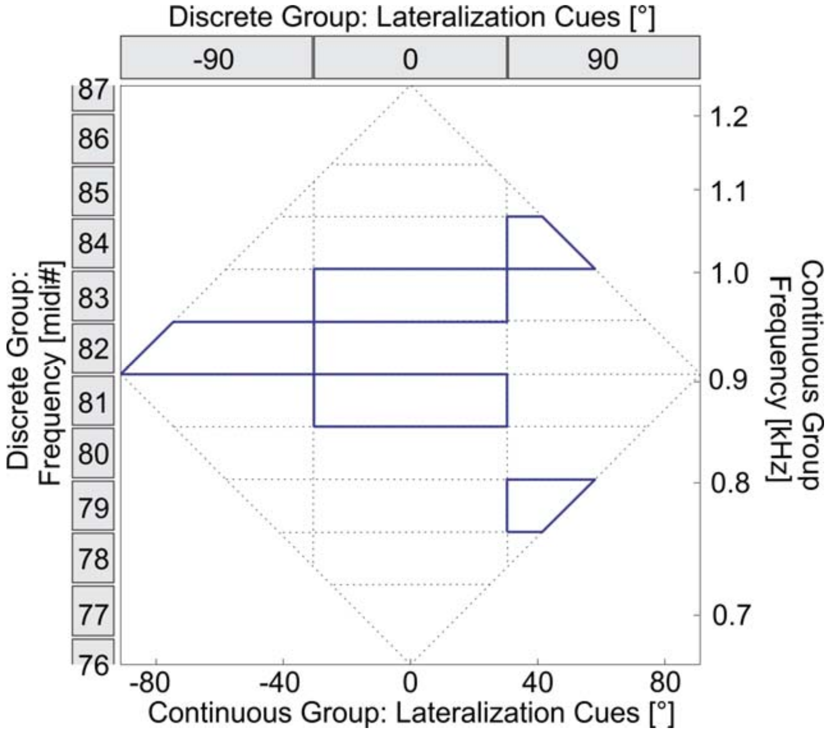

Figure 2. Auditory mapping. Quadrilaterals indicate the five targets participants were asked to achieve. Changes in the difference between two sEMG signals corresponded to spatial cues placing sounds from left to right, either into the discrete categories of -90, 0, and 90 degrees (discrete group, top axis label) or as continuously varying changes from -90 to 90 degrees (continuous group, bottom axis label). Changes in the sum of the two signals corresponded to changes in frequency (perceived pitch), which were either presented as distinct midi notes #76 – 87 (discrete group, left axis label) or as frequency of a tone that varied continuously along the semitone scale (continuous group, right axis label). Because of the way the left and right signals were combined (sum and difference), the area of reachable tone frequency and lateralization had a diamond-like shape.

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, continuous vs. discrete), session (within subjects, days 1 – 3) and the interaction of feedback type × session on ITR. *Post hoc* two-tailed Student t-tests were performed across session (paired), correcting for the multiple (3) comparisons using a conservative Bonferroni correction.

3. Results

ITRs for both continuous and discrete groups were high and improved during the multiple days of training (Figure 3). Individuals in the continuous group reached average ITRs of 15.1 bit/min (SD = 12.8) in session 1, 23.2 bit/min (SD = 25.5) in session 2, and 32.8 bit/min (SD = 27.2) in session 3. Likewise, individuals in the discrete group reached average ITRs of 24.6 bit/min (SD = 14.6) in session 1, 41.3 bit/min (SD = 32.8) in session 2, and 44.3 bit/min (SD = 29.9) in session 3.

Factor	DF	η_p^2	F	*p*
Group	1	0.24	1.53	0.236
Session	2	0.31	6.22	0.006
Group × Session	2	0.02	0.34	0.715

Table 1. Results of ANOVA on information transfer rate

The results of the ANOVA (Table 1) indicated a statically significant effect of session ($p = 0.006$) with a large associated effect size ($\eta_p^2 = 0.31$). Neither group nor the interaction between session and

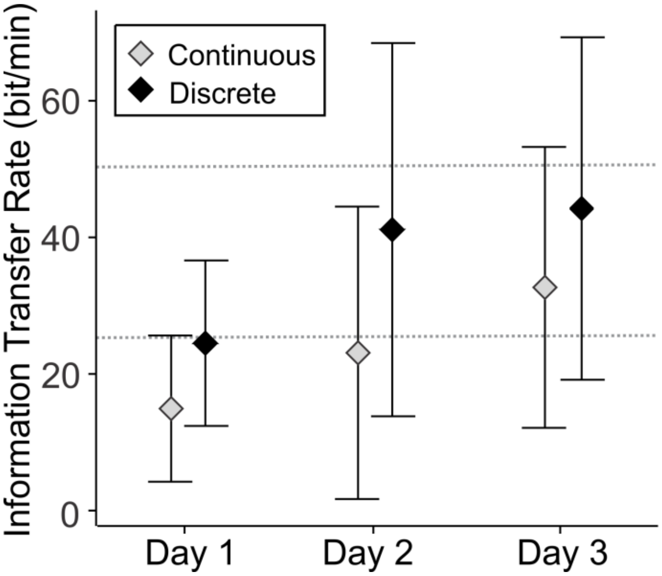

Figure 3. Information transfer rate as a function of group and session. Light grey markers indicate performance of the continuous group; black markers indicate performance of the discrete group. Dotted grey lines indicate performance on similar tasks using surface electromyographic control with *visual* feedback (Williams and Kirsch, 2008; Choi *et al.*, 2011) . *Post-hoc* Tukey's tests determined that ITRs during session 3 were significantly higher relative to sessions 1. Error bars indicate the 95% confidence intervals.

group showed significant effects on ITR. However, there was a trend ($p = 0.24$) for increased ITRs for the discrete group relative to the continuous group. *Post hoc* t-tests across session indicated that ITRs were significantly higher in session 3 relative to session 1 ($p_{\text{adj}} = 0.01$), but no significant differences were seen between session 1 and session 2 ($p_{\text{adj}} = 0.06$) or between session 2 and session 3 ($p_{\text{adj}} = 0.84$).

4. Discussion

This study examined the performance of an HMI using auditory spatial and spectral cues for feedback. We found that individuals in both groups were able to learn from the feedback cues to improve their performance from session 1 to session 3, resulting in increases in ITR over the three sessions.

4.1 Continuous versus discrete feedback

Several prominent theories of motor control contend that a limited subset of variables that are relevant to the specific task are purposefully controlled during movement – allowing variability and instability in task irrelevant dimensions (e.g., Scholz and Schoner, 1999; Todorov and Jordan, 2002). Thus, attention to task-relevant sensory feedback is prioritized and used to inform motor productions, whereas task-irrelevant sensory feedback is ignored. These models have been formulated and tested using examples from limb motor control (utilizing visual and somatosensory feedback), although speech movements also show increased kinematic variability along dimensions that are irrelevant to *auditory* targets (Nieto-Castanon *et al.*, 2005).

Discretization of auditory feedback that is congruent with targets effectively confines the sensory inputs to only those parameters that are task-relevant – that is, continuous motor output is mapped onto discrete outcomes that are directly relevant to the targets in the task space, which may then simplify the processing required by the user. We hypothesized that such purposeful removal of irrelevant feedback would result in improved HMI control performance. Although the results of this experiment only showed a non-significant trend for this improvement, use of the discretized auditory output space (discrete group) resulted in average ITRs that were higher than the continuous group by 63% in session 1, 78% in session 2, and 35% in session 3. More examination in a larger group of participants is necessary to determine whether there are potential effects of discretization on auditory HMI control.

4.2 Performance relative to visual feedback

Both the continuous and the discrete groups were able to achieve high communication rates using purely auditory feedback. After three days of interaction with the auditory HMI, individuals in the two groups were able to attain average ITRs of 38.5 bit/min. This high communication rate is comparable, or even superior, to ITRs from EMG-controlled HMIs utilizing visual feedback for similar tasks, which have reported values ranging from 25.8 – 51.0 bit/min (Williams and Kirsch, 2008; Choi *et al.*, 2011). Future work should directly compare HMI control performance using the auditory feedback developed with visual feedback.

Despite participants' high *average* performance, the large standard deviations (Fig. 3) indicate that there was significant variability in the ability of participants to perform the task. In the current study, musical training was not explicitly controlled for, however post hoc analyses indicated that the two groups had similar distributions of musical training (ranging from novice to expert). Given the potential for musical training to influence performance at this task, future work will determine the effects of musical training on both the continuous and discrete conditions of this task.

5. Conclusions

The goal of this study was to determine the ability of individuals to utilize discrete and continuous pitch and spatial lateralization cues via continuous auditory feedback to control an HMI. ITRs for both continuous and discrete groups were high and improved during the multiple days of training reaching levels consistent with previous HMIs utilizing visual feedback. Although no significant difference between groups was found, there was a trend for higher performance in the group receiving discrete feedback. Future research should explore the potential effects of discrete vs. continuous feedback and level of musical training in a larger group of participants and directly compare the HMI performance outcomes with HMI control using visual feedback.

Acknowledgements

The authors thank Gabrielle Hands for assisting with participant recordings. Portions of this work were funded by NIH NIDCD grant R03DC012651 and a Boston University grant for undergraduate teaching and scholarship.

References

Choi, C., Rim, B. C., and Kim, J. (**2011**). "Development and Evaluation of a Assistive Computer Interface by sEMG for Individuals with Spinal Cord Injuries," in *IEEE International Conference on Rehabilitation Robotics* (Zurich), pp. 1-5.

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

Donchin, E. W. S. a. E. (**Mar 2006**). "A P300-based brain-computer interface: initial tests by ALS patients," Clin Neurophysiol **117**, 538-548.

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

Hermens, H. J., Freriks, B., Merletti, R., Stegeman, D. F., Block, J. H., Gunter, R., ...Hagg, G. M. (**1999**). "European Recommendations for Surface Electromyography: Results of the Seniam Project," (Roessingh Research and Development, Enschede, the Netherlands).

M. Schreuder, B. B., and M. Tangermann (**2010**). "A New Auditory Multi-Class Brain-Computer Interface Paradigm: Spatial Hearing as an Informative Cue," PloS ONE **5**.

Nieto-Castanon, A., Guenther, F. H., Perkell, J. S., and Curtin, H. D. (**2005**). "A modeling investigation of articulatory variability and acoustic stability during American English /r/ production," J Acoust Soc Am **117**, 3196-3212.

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

Pham, M., Hinterberger, T., Neumann, N., Kubler, A., Hofmayer, N., Grether, A., Wilhelm, B., Vatine, J. J., and Birbaumer, N. (**2005**). "An auditory brain-computer interface based on the self-regulation of slow cortical potentials," Neurorehabil Neural Repair **19**, 206-218.
S. P. Kelly, E. C. L., C. Finucane, G. McDarby, and R. B. Reilly (**Sep 2005**). "Visual spatial attention control in an independent brain-computer interface," IEEE Trans Biomed Eng **52**, 1588-1596.
Scavone, G., and Cook, P. (**2005**). "RtMIDI, RtAudio, and a Synthesis (STK) Update,," in Proceedings of the International Computer Music Conference, Barcelona, Spain.
Scholz, J. P., and Schoner, G. (**1999**). "The uncontrolled manifold concept: identifying control variables for a functional task," Exp Brain Res **126**, 289-306.
Todorov, E., and Jordan, M. I. (**2002**). "Optimal feedback control as a theory of motor coordination," Nat Neurosci **5**, 1226-1235.
Williams, M. R., and Kirsch, R. F. (**2008**). "Evaluation of head orientation and neck muscle EMG signals as command inputs to a human-computer interface for individuals with high tetraplegia," IEEE Transactions on Neural Systems and Rehabilitation Engineering **16**, 485-496.