2010 International Conference on Pattern Recognition

Recognizing Sign Language from Brain Imaging

Nishant A. Mehta, Thad Starner,
Melody M. Jackson, Karolyn O. Babalola
Georgia Institute of Technology
Atlanta, GA, USA
{niche,thad,melody}@cc.gatech.edu
karolyn@gatech.edu

G. Andrew James
Dept of Psychiatry and Behavioral Sciences
University of Arkansas for Medical Sciences
Little Rock, AR, USA
gajames@uams.edu

***Abstract*—Classification of complex motor activities from brain imaging is relatively new in the fields of neuroscience and brain-computer interfaces (BCIs). We report sign language classification results for a set of three contrasting pairs of signs. Executed sign accuracy was 93.3%, and imagined sign accuracy was 76.7%. For a full multiclass problem, we used a decision directed acyclic graph of pairwise support vector machines, resulting in 63.3% accuracy for executed sign and 31.4% accuracy for imagined sign. Pairwise comparison of phrases composed of these signs yielded a mean accuracy of 73.4%. These results suggest the possibility of BCIs based on sign language.**

I. Introduction

Nearly two million people in the U.S. suffer from motor disabilities so severe that they cannot communicate [1], [2]. Amyotrophic lateral sclerosis (ALS), a degenerative disease that can gradually destroy motor ability completely, has an estimated global incidence rate near $200,000$ cases per year [3]. Fortunately, brain-computer interfaces (BCIs) can provide new pathways for interaction by using machine learning and signal processing to recognize small changes in brain activity, potentially offering alternate communication channels to the movement-impaired population. We explore the potential of differentiating brain signal patterns associated with motor activity while attempting American Sign Language (ASL). Ultimately, if the effort proves successful, upon diagnosis an ALS patient could begin learning ASL so that when fully locked-in, they can still communicate by attempting to sign.

Compared to the 180 words/minute [4] or 1440 bits/minute[1] of natural languages such as spoken English and ASL, the fastest known BCI today is far slower at 84.7 bits/minute [6]. Our results show the feasibility of recognizing executed and imagined sign from functional magnetic resonance imaging (fMRI) of the brain, demonstrating that a sign-based BCI may be able to boost the bitrate of current BCIs significantly.

[1]A tri-gram model on a one million word corpus of $\sim 25,000$ distinct words yields a perplexity of 247 [5] for written English, yielding an upper bound of ~ 8 bits/word for spoken English.

Previously, researchers have focused on identifying anatomical regions of activation during execution and passive viewing of sign [7]; in contrast, we present results on classifying signs from brain imaging during executed sign and imagined sign. Mitchell et al. [8] performed semantics-driven pairwise classification of brain activity into nouns. Our work includes both verbs and nouns, emphasizes the role of motor cortex, and provides results for multiclass classification problems. Additionally, whereas Mitchell et al. acquired a full brain scan once per second, we acquire three motor cortex slices five times per second. This rapid acquisition may facilitate temporal analysis for sign phrase recognition.

The feasibility of recognizing sign from fMRI draws from work by Rao et al.[9] which suggests that it is possible to resolve individual motor movements with spatially-proximate neural activation if their occurrence is separated by 4 seconds. Additionally, previous works [10], [11], [12] have shown that, for able-bodied subjects, imagined movement produces neural activations similar, though lesser in magnitude, to executed movement.

II. Experiment

We evaluated sign classification ability using our methods on one healthy, right-handed non-native ASL signer. In all experiments, each MRI volume consisted of 3 oblique (axial) slices, covering primary motor cortex (M1), supplementary motor area (SMA), and part of parietal cortex. The first experiment included executed and imagined sign and consisted of an event-related paradigm where, upon visual cue, the subject executed or imagined executing 1 of 7 signs during each trial. For each experiment, the sign cues were presented in an order determined by 5 random permutations of the set of 7 signs. Thus, each experiment included 5 repetitions of each sign cue. The signs, illustrated in Figure 1, correspond to the English words *BED*, *CHAIR*, *COLD*, *HOT*, *I*, *OK*, and *PAIN*. For the purpose of feature selection, we also conducted a block paradigm experiment, with one imagined and one executed block per sign. During the block for a particular task, the task was performed once every 4 seconds for 34 repetitions.

1051-4651/10 $26.00 © 2010 IEEE
DOI 10.1109/ICPR.2010.936

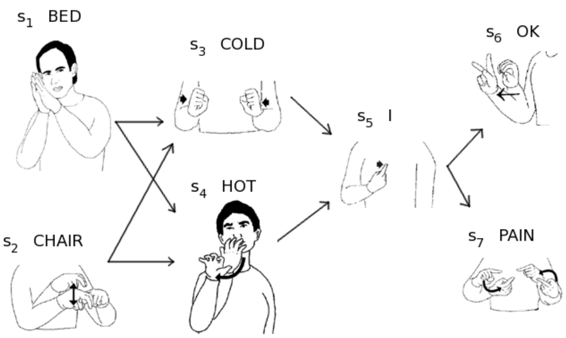

Figure 1. Contrasting sign pairs, *I*, and phrase grammar.

We also acquired event-related data of the subject executing sign phrases. Our set of sign phrases consisted of all eight possibilities of the sequence (*BED* or *CHAIR*), (*COLD* or *HOT*), (*I*), (*OK* or *PAIN*). These phrases were modeled after concepts a locked-in subject might wish to communicate (e.g., *BED HOT I PAIN*). We abbreviate the sequences using the first letter of each sign (e.g. *BED COLD I OK* $\rightarrow$ BCIO). The sign phrases were presented in a random permutation, with 10 repetitions of each sign phrase presented.

III. Methods and Results

Methods.: All the fMRI data was preprocessed using AFNI [13]. We explored multiple classification schemes to test the separability of executed and imagined single sign in the fMRI data. Each of the 3 acquired slices consisted of 64 by 64 voxels, yielding over $12,000$ features. For each trial, we constructed a representative vector by taking the mean activation of all the volumes acquired in the period from 4 to 8 seconds following presentation of the start stimulus.

We used a filtering method for dimensionality reduction so that a classifier could generalize from small training datasets. Consider a classification problem where our goal is to label correctly test points from a class in set S. For each class in S, we take the top k features as ranked by statistical significance of the class versus the rest condition; these significance values are precisely the t-statistics from a general linear model (GLM) (estimated from the block paradigm data) of the activation of each condition (i.e., sign or class). We then use the union of these sets of k features as the feature space. We decompose the multiclass problem into pairwise problems, such that $|S| = 2$.

In our experiments, we tested various methods, such as nearest neighbor classification based on Euclidean distance and cosine similarity, matching examples to prototypes by cosine similarity, generative classifiers based on linear dynamic systems and hidden Markov models, and linear and quadratic kernel support vector machines (SVMs). The performance of linear and quadratic kernel SVMs was far better than the other methods; we present results for linear SVMs because they proved to be far less sensitive to the amount of regularization and the number of features per class k.

A soft-margin linear SVM [14] has the objective

$$\underset{\mathbf{w},b,\boldsymbol{\xi}}{\text{minimize}} \ \|\mathbf{w}\|_2^2 + \lambda \sum_{i=1}^{m} \xi_i \tag{1}$$

$$\text{subject to} \quad y_i(\langle \mathbf{w}, \mathbf{x}_i \rangle + b) \geq 1 - \xi_i, \ \text{for } i = 1, \ldots, m$$
$$\xi_i \geq 0, \ \text{for } i = 1, \ldots, m,$$

where m is the number of training points, x_i is a point in $\mathbf{R}^n$ with label $y_i \in \{-1, 1\}$, λ is used for regularization, and $\boldsymbol{\xi}$ represent the softness of the margin.

For each pair of signs, we trained an SVM classifier using LIBSVM [15]. We set C (see Chang and Lin [16] for the meaning of regularization parameter C) to 100 in all our experiments, and we set k to 60 for executed sign and 15 for imagined sign. The results were not highly sensitive to the particular values of C and k chosen. For the multiclass setting, we used the binary classifiers in a knockout tournament, leading to $l-1$ classification tasks for l classes; this method is a decision directed acyclic graph [17]. We reused the parameters of each binary classifier for the multiclass setting.

Single sign results.: We report all results using leave-one-out cross-validation (LOOCV). Table I presents pairwise classification results for executed and imagined sign. For the sign pairs in Figure 1, mean accuracy was 93.3% for executed and 76.7% for imagined sign.

Table I
Executed (red) and imagined (blue) correctly recognized signs out of 10 pairwise choices.

	Bed		Chair		Cold		Hot		I		OK	
Pain	10	5	9	3	6	6	10	7	10	9	**10**	**9**
OK	6	2	10	6	10	10	9	4	3	6		
I	4	4	10	6	10	9	8	6				
Hot	7	2	8	6	**8**	**8**						
Cold	10	9	7	8								
Chair	**10**	**6**										

Encouraged by these results, we decided to compare all possible pairs, resulting in 83.3% mean accuracy for executed sign and 62.4% accuracy for imagined sign. For executed sign, the pairs for which the classifier had the most difficulty were those whose underlying movement is the most similar. Note that *BED* and *I* both involve right arm movement with brief hand activation for *BED* and brief index finger activation for *I*. Also, *PAIN* and *COLD* are bimanual with brief index finger movement and continuous wrist movement for *PAIN* and brief hand movement and continuous arm movement for *COLD*. *I* and *OK* were not separable, possibly since both involve index finger and arm movement.

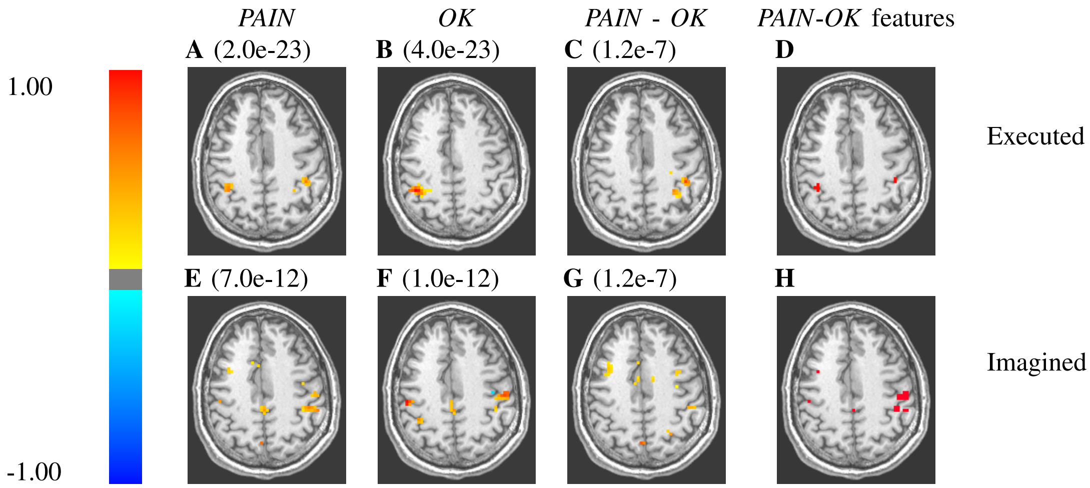

Figure 2. Axial slices (including finger/hand motor areas) of executed and imagined sign.

In Figure 2, we show neural activation for the contrasting pair *PAIN* and OK. From left to right are statistical activations (p-value thresholds in parentheses) for *PAIN*, *OK*, *PAIN* minus *OK*, and at far right features used by the SVM when classifying *PAIN* versus *OK*. In **(A)**, we see that bilateral index finger, thumb, and wrist muscle activity for Executed *PAIN* (EP) produce bilateral activation. In **(B)**, right-handed index finger and thumb opposition in Executed *OK* (EO) produces significant activity in left motor cortex. From **(C)**, we see that the contrast betwen EP and EO is most significant in right motor cortex. In **(D)**, the most significant features used by the SVM classifier for EP versus EO include voxels from both left and right motor cortex.

We also show activations and features for imagined sign. In **(E)**, motor imagery for Imagined *PAIN* (IP) is significant in both left (not shown in this axial slice) and right (shown in this axial slice) motor cortex. As seen in **(F)**, Imagined *OK* (IO) shows significant activation in both motor cortices with stronger activation in left motor cortex. In **(G)**, the contrast between IP and IO indicates widespread differences, including right motor cortex, parietal cortex, and possibly Broca's area (Brodmann areas 44 and 45). As expected, **(H)** shows that the most significant features for the SVM classifier of IP versus IO concentrate primarily in right motor cortex.

For the multiclass setting (see Figure 3), the classifier was forced to choose one sign out of all seven. Executed sign yielded 51.4% accuracy and imagined sign yielded 31.4% accuracy. Note that the chance level when choosing from a set of seven signs is 14.3%. Removing *I* (the simplest sign and only included for phrase composition) for executed sign resulted in 63.3% accuracy.

Note that the contrasting pairs of signs are bimanual versus unimanual. We explored the possibility that the classifier performs no better than a baseline classifier which only discriminates bimanual versus unimanual signs. This baseline classifier's expected accuracy is 33.3% (3 unimanual and 3 bimanual signs), still well below the accuracy realized for our executed sign classifier; however, this accuracy is nearly the same as our imagined sign classifier. Even so, attempted motion by the motor impaired produces neural activation more similar to that of executed motion in able-bodied subjects than imagined motion [12], providing hope that finer sign discrimination is possible with locked-in subjects.

Sign phrase results.: For sign phrase feature selection, using only fMRI data from the sign phrase session, we computed the statistical significance of each voxel in determining whether the subject is executing any sign phrase or no sign phrase; we used all the data but no labels. We selected the 100 most significant voxels from the GLM, yielding a feature set that generally is relevant to all the sign phrases. To embed a sign phrase trial as a point, we averaged the data from 5 to 12 seconds following the presentation of the start stimulus for that trial, and we spatially restricted the features to the 100 most significant voxels. We performed LOOCV using the same SVM as in the single sign case (with $C = 0.1$) for all pairs of the 8 sign phrases (10 examples per phrase). The mean accuracy across all 28 binary classification tasks was 73.4% (see Figure 4).

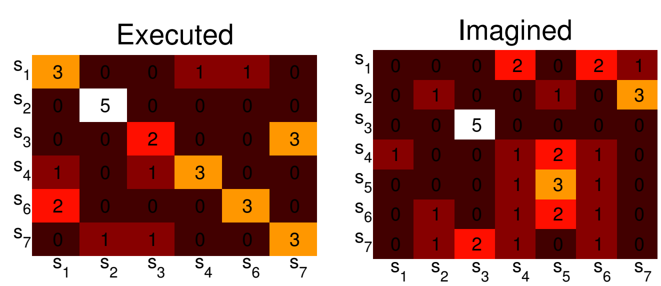

Figure 3. Multiclass confusion matrices.

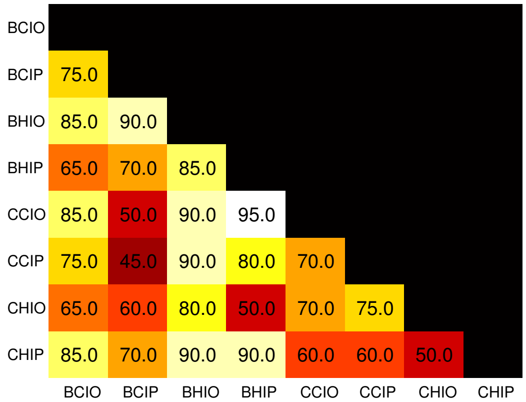

Figure 4. Accuracy (in %) for all pairs of sign phrase classification tasks.

IV. Conclusion

We have presented results suggesting the feasibility of sign recognition using brain images. Our focus now is to explore temporal statistical models of sign phrase activation using models of single sign activation. If we can successfully perform multiclass classification on a larger set of sign phrases from fMRI, the next step is to determine whether attempted sign phrases can be classified from recordings via a portable modality such as electroencephalography. With proper vocabulary selection and the development of a more powerful sign phrase classifier, we hope to narrow the gap between the bitrates of BCIs and natural language.

Acknowledgements

We thank Paul Corballis for his expertise with fMRI experiment design. We also would like to thank the late Randy Pausch, whose constructive criticism led to our current approaches. This research was funded by the National Science Foundation under award IIS-0836747.

References

[1] R. C. Ficke, N. I. on Disability, R. R. (US), and S. M. Corporation, *Digest of Data on Persons with Disabilities*. National Institute on Disability and Rehabilitation Research, 1992.

[2] National Advisory Board on Medical Rehabilitation Research, *Report and Research Plan for the National Center for Medical Rehabilitation Research*. NICHHD, NIH, 1992.

[3] P. Preux, M. Druet-Cabanac, P. Couratier, C. Debrock, T. Truong, W. Marcharia, J. Vallat, M. Dumas, and F. Boutros-Toni, "Estimation of the amyotrophic lateral sclerosis incidence by capture-recapture method in the Limousin region of France," *Journal of Clinical Epidemiology*, vol. 53, no. 10, pp. 1025–1029, 2000.

[4] J. J. Darragh and I. H. Witten, "The reactive keyboard. cambridge series on human-computer interaction," 1992.

[5] R. A. Cole, J. Mariani, H. Uszkoreit, A. Zaenen, V. Zue *et al.*, *Survey of the state of the art in human language technology*. Cambridge University Press, 1997.

[6] P. Meinicke, M. Kaper, F. Hoppe, M. Heumann, and H. Ritter, "Improving transfer rates in brain computer interfacing: A case study," *Neural Information Processing Systems*, pp. 1131–1138, 2003.

[7] L. San Jose-Robertson, D. Corina, D. Ackerman, A. Guillemin, and A. Braun, "Neural systems for sign language production: Mechanisms supporting lexical selection, phonological encoding, and articulation," *Human Brain Mapping*, vol. 23, no. 3, 2004.

[8] T. M. Mitchell, S. V. Shinkareva, A. Carlson, K. M. Chang, V. L. Malave, R. A. Mason, and M. A. Just, "Predicting human brain activity associated with the meanings of nouns," *Science*, vol. 320, no. 5880, p. 1191, 2008.

[9] S. M. Rao, P. A. Bandettini, J. R. Binder, J. A. Bobholz, T. A. Hammeke, E. A. Stein, and J. S. Hyde, "Relationship between finger movement rate and functional magnetic resonance signal change in human primary motor cortex," *Journal of Cerebral Blood Flow & Metab.*, vol. 16, 1996.

[10] R. Beisteiner, P. Hollinger, G. Lindinger, W. Lang, and A. Berthoz, "Mental representations of movements. Brain potentials associated with imagination of hand movements," *Electroencephalography and Clinical Neurophysiology*, vol. 96, no. 2, pp. 183–93, 1995.

[11] M. Lotze *et al.*, "Activation of cortical and cerebellar motor areas during executed and imagined hand movements: An fMRI study," *Journal of Cognitive Neuroscience*, vol. 11, no. 5, pp. 491–501, 1999.

[12] H. Alkadhi, P. Brugger, S. H. Boendermaker, G. Crelier, A. Curt, M. C. Hepp-Reymond, and S. S. Kollias, "What disconnection tells about motor imagery: evidence from paraplegic patients," *Cerebral Cortex*, vol. 15, no. 2, pp. 131–140, 2005.

[13] R. Cox, "AFNI: software for analysis and visualization of functional magnetic resonance neuroimages," *Computers and Biomedical Research*, vol. 29, no. 3, pp. 162–173, 1996.

[14] B. Schölkopf and A. J. Smola, *Learning with kernels*. MIT press Cambridge, Mass, 2002.

[15] C.-C. Chang and C.-J. Lin, *LIBSVM: a library for support vector machines*, 2001, software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.

[16] R. Fan, P. Chen, and C. Lin, "Working set selection using second order information for training support vector machines," *Journal of Machine Learning Research*, vol. 6, p. 1918, 2005.

[17] J. Platt, N. Cristianini, and J. Shawe-Taylor, "Large margin DAGs for multiclass classification," *Neural Information Processing Systems*, vol. 12, pp. 547–553, 2000.