Recognizing Sign Language from Brain Imaging

Nishant A. Mehta[1*], Thad Starner[1], Melody Moore Jackson[1],
Karolyn O. Babalola[2], George Andrew James[3]

[1]College of Computing, Georgia Institute of Technology,
Technology Square Research Building, 85 5th St NW, Atlanta, GA 30308, USA

[2]School of Electrical and Computer Engineering, Georgia Institute of Technology,
777 Atlantic Drive NW, Atlanta, GA 30332-0250, USA

[3]Biomedical Imaging Technology Center, Department of Biomedical Engineering,
Georgia Institute of Technology/Emory University,
531 Asbury Circle, Suite N305, Atlanta, GA 30322-4600, USA

[*]To whom correspondence should be addressed; E-mail: niche@cc.gatech.edu.

Abstract

The problem of classifying complex motor activities from brain imaging is rel-
atively new territory within the fields of neuroscience and brain-computer inter-
faces. We report positive sign language classification results using a tournament
of pairwise support vector machine classifiers for a set of 6 executed signs and
also for a set of 6 imagined signs. For a set of 3 contrasted pairs of signs, executed
sign and imagined sign classification accuracies were highly significant at 96.7%
and 73.3% respectively. Multiclass classification results also were highly signif-
icant at 66.7% for executed sign and 50% for imagined sign. These results lay
the groundwork for a brain-computer interface based on imagined sign language,
with the potential to enable communication in the nearly 200,000 individuals that
develop progressive muscular diseases each year.

1 Introduction

Nearly two million people in the U.S. alone suffer from motor disabilities so severe that they cannot
communicate [1, 2, 3, 4]. Amyotrophic lateral sclerosis (ALS), which is an example of a degen-
erative disease that can gradually destroy motor ability completely, has an estimated worldwide
incidence rate near $200,000$ cases per year [5]. Brain-computer interfaces (BCIs) can provide new
pathways for interaction by using machine learning and signal processing methods to recognize
small changes in brain activity, potentially offering alternate channels of communication to the
movement-impaired population. This paper explores the potential of differentiating brain signal
patterns associated with imagined motor movements in the form of American Sign Language.

Compared to the 180 words/minute [6] or 1440 bits/minute[1] of natural languages such as spoken En-
glish and American Sign Language, the fastest known BCI today is far slower at 84.7 bits/minute [8].
We report results showing that it is possible to recognize both executed and imagined sign language
from functional magnetic resonance imaging (fMRI) of the brain, demonstrating the possibility that
a sign-based BCI can boost the communication rate of current BCIs to that of natural language users.
Previously, researchers focused on identifying anatomical regions of activation during execution and
passive viewing of sign [9]; in contrast, we present results on classifying signs from brain imaging
during executed sign and imagined sign. Mitchell et al. [10] performed semantics-driven pairwise

---

[1]Calculations using a tri-gram model on a one million word corpus of $20,000$ - $30,000$ distinct words yield
a perplexity of 247 [7] for written English, upper bounding the perplexity of spoken English. This translates to
approximately 8 bits/word, or 1440 bits/minute.

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 can facilitate temporal analysis for extensions to sign phrase recognition.

The feasibility of recognizing sign from fMRI draws from work by Rao et al. [11] which suggests that it is possible to resolve individual motor movements with spatially-proximate neural activation if their occurrence is separated by 4 seconds. Regarding imagined sign classification, previous works [12, 13, 14, 15, 16] have shown that, for able-bodied subjects, imagined movement produces neural activations similar, though lesser in magnitude, to executed movement.

Our emphasis is to create assistive devices for individuals with progressive muscular diseases such as ALS. After being diagnosed with ALS and during the disease's progression, a patient can learn a useful sign vocabulary. When ultimately incapable of movement (i.e. "locked-in"), attempts to produce the signs will not result in physical movement but can create the signs' associated activation patterns in motor cortex. A BCI can then recognize the patterns associated with these attempted signs and translate them to language.

Sign languages appear more attractive than vocal speech from a brain imaging perspective. Vocal speech involves muscles in a relatively small region of the motor cortex [17], and phonemes are relatively indistinct from each other at the current brain imaging spatiotemporal resolutions. Signs, however, are constructed using coarser muscle movements and often vastly differ from each other: one-versus-two handed execution, static-versus-sweeping arm postures, fixed-versus-dynamic finger poses, and symmetry of left and right hand shapes [18].

2 Experiments

We evaluated sign classification ability using our methods on one healthy, right-handed non-native signer of American Sign Language. We conducted all experiments with both executed and imagined sign. The experiments 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*. These signs were chosen for their relevance to the motor-impaired. 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, visual cue for task performance occurred once every 4 seconds for 34 repetitions. All recordings were done using a 3T MR scanner (Magnetom Trio; Siemens) with a volume acquisition repetition time of 200 ms (TR/TE/FA = 200ms/28.3ms/90°; FOV 192x192mm; 3 oblique slices 4mm thick each; matrix size=64x64; final voxel resolution 3x3x4mm^3). Each volume consisted of 3 oblique slices[19], covering primary motor cortex (M1), supplementary motor area (SMA), and part of parietal cortex.

3 Methods and Results

We explored multiple classification schemes to test the separability of executed and imagined sign in the fMRI data. Each of the 3 acquired slices consisted of 64 by 64 voxels yielding an acquired volume of over $12,000$ dimensions. For each trial, we constructed a representative vector by taking the mean activation from 4 to 8 seconds following presentation of the start stimulus for that trial. Due to the data's high dimensionality, we pursued dimensionality reduction so that a classifier could generalize from small training datasets. To this end, we applied a filtering method for feature selection which, for a classification task with class set S, selects the union of the k features whose activation (versus the rest condition) is most statistically significant for each class in S.

In our experiments, classifiers that directly focus on modeling the class boundaries performed best. These classifiers include both linear and quadratic support vector machines (SVMs) [20], which
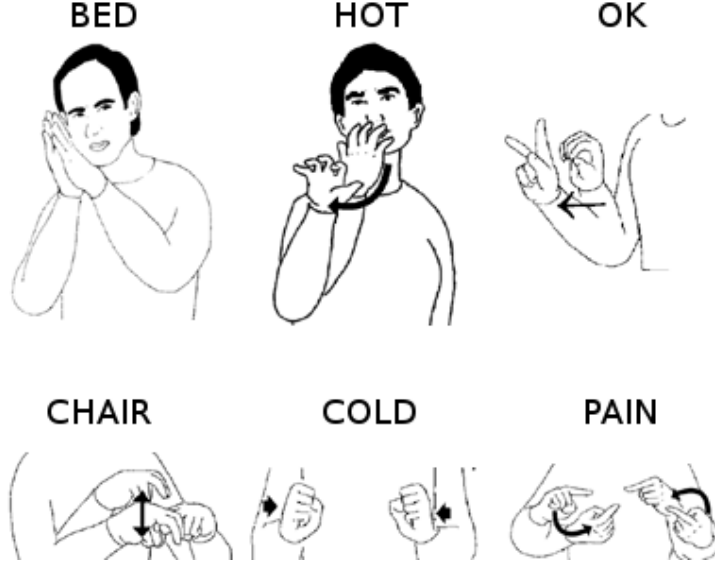
Figure 1: American Sign Language signs used in experiments, chosen for use in forced-choice questions with motor-impaired patients, such as "Do you want to go to your *BED* or *CHAIR*?," "Are you feeling *COLD* or *HOT*?," and "Are you in *PAIN* or *OK*?." The sign *I* can be used to construct sign phrases. We modified *BED* such that only the right hand moves.

minimize the following objective:

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

$$\text{subject to} \ \ y_i(\langle w, \phi(x_i)\rangle + b) \geq 1 - \xi_i, \ \text{for} \ i = 1, 2, \ldots, m,$$

where m is the number of training points, x_i is a point in $\mathbf{R}^n$ with label $y_i \in \{-1, 1\}$, the variable C and variables ξ_i are used for regularization when the classes are not perfectly separable, and ϕ is the feature map $\phi : \mathbf{R}^n \mapsto \mathcal{H}$, for a Hilbert space $\mathcal{H}$. For the linear SVM, ϕ is the identity mapping, while for the quadratic SVM, if $x_i = (x_{i,1}, \ x_{i,2})$, then

$$\phi(x_i) = (x_{i,1}^2, \ \sqrt{2}x_{i,1}x_{i,2}, \ x_{i,2}^2, \ \sqrt{2}x_{i,1}, \ \sqrt{2}x_{i,2}, \ 1),$$

capturing as features all monomials of order two and lower. The SVM problem is convex and can be solved efficiently [20].

We briefly describe the idea behind the problem. Assume we have observations $Z = (X, Y)$, where $X = (x_1, x_2, \ldots, x_m)$ such that $x_i \in \mathbf{R}^n$, and $Y = (y_1, y_2, \ldots, y_m)$ for $y_i \in \{-1, 1\}$, $1 \leq i \leq m$. If the negatively and positively labeled points are separable perfectly[2], an SVM finds a separating hyperplane with normal vector w which maximizes the margin, the minimum distance of the hyperplane to points of either class. It has been shown that minimizing $\|w\|^2$, which is equivalent to maximizing the margin over a training set of points, both theoretically [21] and empirically [22] leads to lower error over a test set of points.

Increasing the margin decreases an upper bound on the Vapnik-Chervonenkis (VC) dimension of the classification algorithm; it can be shown that the test error on new data can be upper bounded by the sum of the training error and a term that decreases with the VC dimension [20]. Hence, up to a point, preferring classifers with low VC dimension (max margin, low $\|w\|^2$) provides tighter bounds on the test error. As a result, decreasing the dimensionality of w (effectively reducing $\|w\|^2$) theoretically can reduce the test error, justifying some feature selection.

---

[2]The points may not be separable in the feature space, in which case we augment the SVM objective with a weight on the margin and another term which relates to the softness of the margin (the amount by which examples can lie on the wrong side of the separating hyperplane). The softness is reflected in the ξ_i's and the regularization parameter C. See Schölkopf and Smola [20] or Burges [21] for more depth.

	BED		CHAIR		COLD		HOT		I		OK	
PAIN	100	70*	90	80	70	80	100	50	100	90	*100*	*90*
OK	80	80*	100	80*	100	90	90	70*	60	70*		
I	70*	80*	100	100*	100	80	90	80*				
HOT	80	50*	100	60*	*90*	*70**						
COLD	100	80	90	50								
CHAIR	*100*	*60**										

Table 1: Executed sign (red background) and imagined sign (blue background) pairwise classification results computed using leave-one-out cross-validation with 5 examples of each sign (10 examples per pairwise test). Italicized entries indicate contrasted pairs. The mean accuracy on the contrasted pairs is 96.7% for executed sign and 73.3% for imagined sign. Non-asterisked results are from linear SVMS. Asterisks indicate results from quadratic SVMs. Note that for each pairwise task, 80% accuracy indicates statistically significant class separation with greater than 90% confidence, 70% accuracy is weakly significant with greater than 82% confidence, and 60% accuracy is not significant.

For each pair of signs, we trained a binary SVM classifier. We selected the best settings for the regularization parameter C and the number of filtered features k, first using a linear SVM and then a quadratic SVM if the linear classifier was not able to separate the classes well. Assume the observations $Z = (X, Y)$, where $X = (x_1, x_2, \ldots, x_m)$ such that $x_i \in \mathbf{R}^n$, and $Y = (y_1, y_2, \ldots, y_m)$ for $y_i \in \{-1, 1\}, 1 \le i \le m$. The criterion for parameter selection was leave-one-out cross-validation, where for a particular parameter configuration we test the classifier on each point x_i by training on a leave-one-out set of points and labels $Z_{-i} = Z \setminus \{z_i\}$. The results were not highly sensitive to the particular values of the parameters chosen. For the multiclass setting, we use the binary classifiers in a knockout tournament, leading to $l - 1$ classification tasks for l classes. This method is known as a decision directed acyclic graph [23]. Each binary classifier used in the multiclass setting for a particular mode (executed versus imagined) takes on the same parameter configuration as that binary classifier used in the pairwise classification setting for that mode.

In our experiments we found the above technique to provide better results than a cosine-based classifier used by Mitchell et al. [10]. Table 1 presents pairwise classification results for executed and imagined sign. For executed sign, the pairs for which the classifier had the most difficulty are 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* both 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 (see Table 1), which may be explained by both signs involving index finger movement and arm movement.

Multiclass results using all signs yielded 48.6% for the executed case and 40% accuracy for the imagined case, well above the 14.3% chance level when classifying 7 signs. For reasons explained below, we removed *I* from the executed case and *HOT* from the imagined case, yielding 66.7% accuracy and 50% accuracy respectively. These results for executed and imagined sign are presented in Table 2. The sign *I* was removed from the executed sign multiclass task due to poor separability from the other signs; this is no great loss because the sign *I* was included solely for future work with sign phrases. The executed sign accuracy (after removing *I*) of 66.7% is significantly greater than the chance level of 16.7% for 6 signs. We also explored the possibility that the classifier performs no better than a baseline classifier which only discriminates bimanual versus right unimanual signs. This baseline classifier's expected accuracy is 33.3% (3 unimanual signs, 3 bimanual signs), still well below the accuracy realized for our classifier. Certain patterns emerge when analyzing the classification errors. When the ground truth was a unimanual sign (*BED*, *HOT*, *OK*), erroneous predictions often also were unimanual signs. When the ground truth was *COLD*, both erroneous predictions were *PAIN*, which in our set is the most similar sign to *COLD*. Erroneous predictions when the ground truth was *PAIN* also were other bimanual signs.

The imagined sign results were generally positive, although the results for most pairs containing *HOT* were unexpectedly poor. Upon reflection, we realized that *HOT* is very similar to *BED*, and so

4

true\prediction	BED		CHAIR		COLD		HOT		I		OK		PAIN	
BED	3	2	0	2	0	0	1	X	X	0	1	1	0	0
CHAIR	0	2	5	1	0	1	0	X	X	0	0	0	0	1
COLD	0	0	0	1	3	3	0	X	X	0	0	0	2	1
HOT	1	X	0	X	1	X	3	X	X	X	0	X	0	X
I	X	0	X	0	X	0	X	X	X	4	X	0	X	1
OK	1	0	0	0	0	0	1	X	X	2	3	2	0	1
PAIN	0	0	1	1	1	1	0	X	X	0	0	0	3	3

Table 2: Executed sign (red background) and imagined sign (blue background) multiclass confusion matrix. The overall executed sign accuracy was 66.7%. X's indicate signs not used for a particular multiclass task. The overall imagined sign accuracy was 50%.

we removed *HOT* in the imagined multiclass scenario. In particular, the slight difference between *HOT* and *BED*, involving only the brief wrist flick of *HOT*, appears sufficient for separating their executed versions but not their imagined versions. Surprisingly, Imagined *BED* and *CHAIR* also were not well separable. Although *CHAIR* is bimanual, the role of the non-dominant hand is minimal, suggesting less vivid motor imagery, while the role of the dominant hand for the two signs differs only in the upper digits.

The imagined sign multiclass results also were well above those expected by chance or using the handedness-based classifier. The imagined multiclass accuracy (after removing *HOT*) was 50%, also greater than chance or our baseline classifier. As expected from the pairwise classification results, *BED* and *CHAIR* were not well separable; in the multiclass setting, each was misclassified twice as the other. When the ground truth was each of *COLD* or *PAIN*, both misclassifications were other bimanual signs.
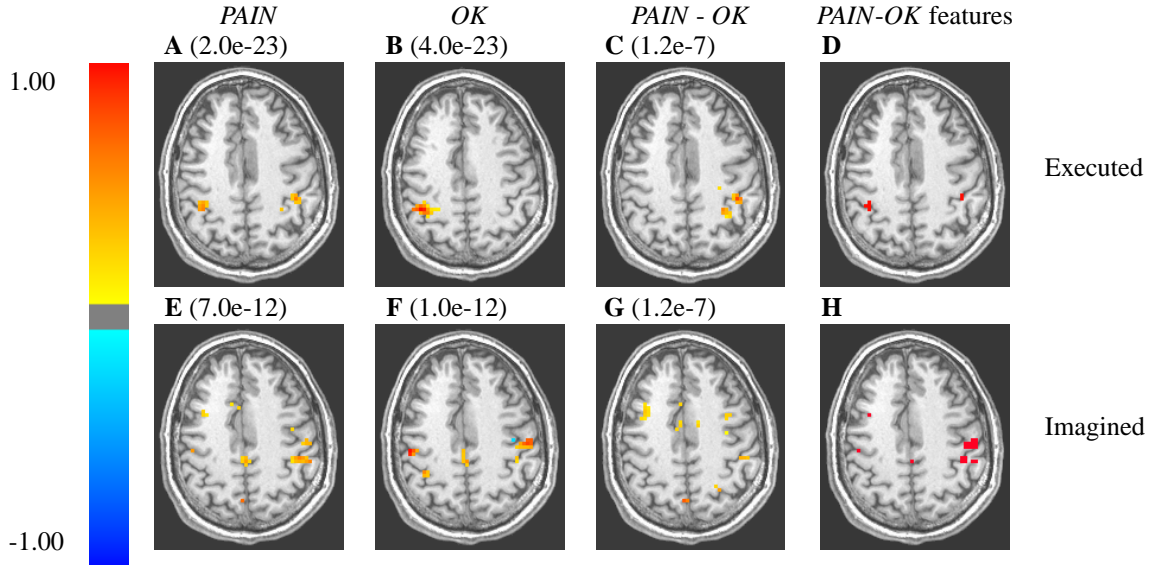
Figure 2: Axial slices of executed and imagined sign. Left to right: Statistical activations (p-value thresholds in parentheses) for *PAIN*, *OK*, *PAIN* minus *OK*, and at far right features used by the SVM for the *PAIN* versus *OK* classification task. **(A)** Bilateral index finger, thumb, and wrist muscle activity for Executed *PAIN* (EP) produce bilateral activation. **(B)** Right-handed index finger and thumb opposition in Executed *OK* (EO) produces significant activity in left motor cortex. **(C)** EP and EO contrast is most significant in right motor cortex. **(D)** SVM classifier for EP versus EO utilizes voxels from both left and right motor cortex. **(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. **(F)** Imagined *OK* (IO) shows significant activation in both motor cortices with stronger activation in left motor cortex. **(G)** Contrast between IP and IO indicates widespread differences, including right motor cortex, parietal cortex, and possibly Broca's area (Brodmann areas 44 and 45). **(H)** Features for SVM classifier of IP versus IO concentrate primarily in right motor cortex.

In Figure 2, we show axial slices relevant to discriminating *PAIN* versus *OK*. The three slices for Executed *PAIN*, Executed *PAIN* minus *OK*, and the features used for Executed *PAIN* versus *OK* classification all are overlaid onto the same anatomical slice. The slice for Executed *OK* is 3mm superior to the 3 slices, while all of the imagined sign slices are 2mm superior to the 3 slices. Executed *PAIN* (bimanual wrist and index finger) shows the expected bilateral activation in M1, while Executed *OK* (right arm, thumb, and index finger) shows the expected activation in left motor cortex with a much larger region of left motor cortex activated than in *PAIN*. Although the contrast for Executed *PAIN* minus *OK* indicates right M1, the classifier for Executed *PAIN* versus *OK* uses many voxels from left M1 as well as several from right M1.

The slice shown for Imagined *PAIN* has significant activation in right M1 and SMA. Another slice, not shown, for Imagined *PAIN* shows significant left M1 activation. The slice for Imagined *OK* shows similar activation in SMA, with the more posterior M1 activation primarily concentrated to the left. The slice for Imagined *PAIN* minus *OK* indicates significant activation primarily in right M1, SMA, and somewhat surprisingly what appears to be Broca's area (left hemisphere). It has been shown [24] that Broca's area is activated during executed signs for bilinguals for which English and American Sign Language are native languages. A possible explanation for differential activation of Broca's area for the two signs may be the additional semantic complexity of *PAIN* as compared to the less informative *OK*; *OK* often is semantically ambiguous/uninformative in spoken English. Finally, the classifier for Imagined *PAIN* versus *OK* uses voxels concentrated mostly in right M1 and SMA, in agreement with the motor differences between the two signs.

4   Discussion.

Numerous researchers have explored differential activation in the brain for executed and imagined movement. An immediate concern here is how well our results will translate to future results with movement-impaired populations. Alkadhi et al. [16] performed a study that investigated neural activation for executed and imagined right foot movement in healthy subjects and imagined foot movement in spinal cord injury (SCI) subjects. While healthy subject motor imagery showed no significant activation in M1 nor primary somatosensory cortex (S1), SCI imagery subjects showed significant activation contralaterally in M1 and bilaterally in S1. Additionally, the M1 activation found for SCI imagery was similar to the M1 activation for healthy movement execution (ME), with no M1 activation remaining in SCI imagery when contrasting in either direction with healthy ME. The implication of this result is that, when movement-impaired individuals imagine sign, we expect M1 activation to be close to that of the executed movement condition. In short, good classification results for executed sign tasks in healthy subjects should support good classification results for imagined sign tasks in motor-disabled subjects.

Given the results attained so far with sign recognition in the brain, we are exploring sign phrase recognition to attempt substantial increases in the communication rate of BCIs. Portability is crucial to the goal of developing a BCI for the movement-impaired. If sign phrase classification from fMRI proves successful, the next step is to determine whether attempted sign phrases can be classified from recordings via a portable modality such as electroencephalography (EEG). For the purpose of sign discrimination, EEG quite beneficially has very fine temporal resolution (on the order of milliseconds), but unfortunately its spatial resolution is quite coarse; however, recent advances in using EEG in combination with spatial information from fMRI data indicate that the spatial resolution of EEG can be boosted substantially.

Mapping scalp-recorded EEG to three-dimensional source locations is an ill-posed problem, but constraining mappings by fMRI prior knowledge greatly makes the problem well-posed. Im and Lee [25] developed a method for weighting fMRI functional priors according to how well the fMRI matches spatial activation inferred from EEG alone. Additional research by Ahflors et al. [26] uses fMRI to find electric dipoles, while Wagner et al. [27] demonstrated fMRI-constrained cortical current density estimation. The most promising support for our work may be the demonstration by Dale et al. [28] that fMRI recorded during stimulus presentation can be incorporated into EEG source localization, using Bayesian statistics and an approximation to electrical activity/hemodynamic response dependency.

Our classification results highlight the importance of selecting a set of maximally discriminative signs. This set can be customized for a particular user's needs and neural activations. For example,

a caregiver could ask "What's wrong?," and the user could respond by attempting to sign *HOT*, *COLD*, *PAIN*, or *OK*. We are expanding the sign vocabulary and may even entirely create new signs, with emphasis on discriminability and utility. With proper vocabulary selection, expansion to sign phrases, and the use of machine learning sequence classification methods, we hope to narrow the gap between the information transmission rates of BCIs and natural language.

References

[1] R. C. Ficke, N. I. on Disability, R. R. (US), 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* (National Institute of Child Health and Human Development, National Institute of Health, 1992).

[3] C. Murray, A. Lopez, *Lancet* **349** (1997).

[4] G. T. Carter, *Journal of Neurologic Rehabilitation* **11**, 69 (1997).

[5] P. M. Preux, *et al.*, *Journal of Clinical Epidemiology* **53**, 1025 (2000).

[6] J. J. Darragh, I. H. Witten, The reactive keyboard. cambridge series on human-computer interaction (1992).

[7] R. A. Cole, *et al.*, *Survey of the state of the art in human language technology* (Cambridge University Press Cambridge, 1997).

[8] P. Meinicke, M. Kaper, F. Hoppe, M. Heumann, H. Ritter, *Advances in Neural Information Processing Systems* pp. 1131–1138 (2003).

[9] L. San Jose-Robertson, D. P. Corina, D. Ackerman, A. Guillemin, A. R. Braun, *Human brain mapping* **23** (2004).

[10] T. M. Mitchell, *et al.*, *Science* **320**, 1191 (2008).

[11] S. M. Rao, *et al.*, *Journal of Cerebral Blood Flow & Metabolism* **16**, 1250 (1996).

[12] R. Beisteiner, P. Hollinger, G. Lindinger, W. Lang, A. Berthoz, *Electroencephalography and Clinical Neurophysiology* **96**, 183 (1995).

[13] G. Pfurtscheller, C. Neuper, *Neuroscience Letters* **239**, 65 (1997).

[14] W. Lang, D. Cheyne, P. Höllinger, W. Gerschlager, G. Lindinger, *Cognitive brain research* **3**, 125 (1996).

[15] M. Lotze, *et al.*, *Journal of Cognitive Neuroscience* **11**, 491 (1999).

[16] H. Alkadhi, *et al.*, *Cerebral Cortex* **15**, 131 (2005).

[17] R. E. Martin, *et al.*, *Journal of neurophysiology* **92**, 2428 (2004).

[18] S. K. Liddell, R. E. Johnson, *Linguistics of American Sign Language: An introduction*, C. Valli, C. Lucas, eds. (Gallaudet University Press, 2001), pp. 267–306, third edn.

[19] Materials and methods are available as supporting material on *Science* Online .

[20] B. Schölkopf, A. J. Smola, *Learning with kernels* (MIT press Cambridge, Mass, 2002).

[21] C. J. C. Burges, *Data mining and knowledge discovery* **2**, 121 (1998).

[22] J. Langford, *Journal of Machine Learning Research* **6**, 273 (2006).

[23] J. C. Platt, N. Cristianini, J. Shawe-Taylor, *Advances in neural information processing systems* **12**, 547 (2000).

[24] B. Horwitz, *et al.*, *Neuropsychologia* **41**, 1868 (2003).

[25] C. H. Im, S. Y. Lee, *Physics in Medicine and Biology* **51**, 6005 (2006).

[26] S. Ahlfors, *et al.*, *Journal of Neurophysiology* **82**, 2545 (1999).

[27] M. Wagner, M. Fuchs, J. Kastner, *Proc 12th Intl Conf Biomagnetism* (2000), pp. 785–788.

[28] A. M. Dale, *et al.*, *Neuron* **26**, 55 (2000).