SCIENTIFIC
REPORTS
natureresearch

OPEN

Optimising non-invasive brain-computer interface systems for free communication between naïve human participants

Angela I. Renton [1*], Jason B. Mattingley [1,2,3] & David R. Painter[2]

Free communication is one of the cornerstones of modern civilisation. While manual keyboards currently allow us to interface with computers and manifest our thoughts, a next frontier is communication without manual input. Brain-computer interface (BCI) spellers often achieve this by decoding patterns of neural activity as users attend to flickering keyboard displays. To date, the highest performing spellers report typing rates of ~10.00 words/minute. While impressive, these rates are typically calculated for experienced users repetitively typing single phrases. It is therefore not clear whether naïve users are able to achieve such high rates with the added cognitive load of genuine free communication, which involves continuously generating and spelling novel words and phrases. In two experiments, we developed an open-source, high-performance, non-invasive BCI speller and examined its feasibility for free communication. The BCI speller required users to focus their visual attention on a flickering keyboard display, thereby producing unique cortical activity patterns for each key, which were decoded using filter-bank canonical correlation analysis. In Experiment 1, we tested whether seventeen naïve users could maintain rapid typing during prompted free word association. We found that information transfer rates were indeed slower during this free communication task than during typing of a cued character sequence. In Experiment 2, we further evaluated the speller's efficacy for free communication by developing a messaging interface, allowing users to engage in free conversation. The results showed that free communication was possible, but that information transfer was reduced by voluntary textual corrections and turn-taking during conversation. We evaluated a number of factors affecting the suitability of BCI spellers for free communication, and make specific recommendations for improving classification accuracy and usability. Overall, we found that developing a BCI speller for free communication requires a focus on usability over reduced character selection time, and as such, future performance appraisals should be based on genuine free communication scenarios.

Overview. Free and open communication is fundamental to modern life, scientific enterprise and democratic discourse. The rising prevalence of technologies such as virtual/augmented reality[1–6] has created new opportunities for hands-free communication and control. Brain-computer interfaces (BCI), which translate measurements of the user's brain activity into computer commands to control external devices[7–11], present emerging forms of hands-free communication. BCI spellers are virtual keyboards that decode brain activity patterns allowing users to select characters in sequence to spell words and, ultimately, freely communicate[12]. BCI keyboards mimic manual keyboards, which extend the user by allowing them to physically manifest their real-time thoughts, interface with the internet and communicate remotely. BCI communication systems, including spellers, have long been used in clinical settings to facilitate communication in cases of quadriplegia, anarthria and amyotrophic lateral sclerosis[13–15]. These systems are often developed using electroencephalography (EEG), which allows for portable, flexible and affordable devices[16]. BCI has the potential to revolutionise communication, and yet its potential for creating free communication in healthy users is largely unexplored[17]. Here, we developed an open-source,

[1]Queensland Brain Institute, The University of Queensland, St Lucia, 4072, Australia. [2]School of Psychology, The University of Queensland, St Lucia, 4072, Australia. [3]Canadian Institute for Advanced Research (CIFAR), Toronto, Canada. *email: angie.renton23@gmail.com

1

high-performance, non-invasive BCI communication system, and explored how the numerous free parameters involved in testing, design, and algorithm implementation can be altered to optimise performance and usability.

Remarkable progress had been made in the development of signal processing and classification algorithms to decode the brain activity underlying BCI control commands[12,18–21], including faster and more accurate classification of neural activity evoked by flickering virtual keyboards[6,12,21,22]. For communication, the original and most prevalent BCI speller uses the P300, an event-related potential (ERP) component modulated by attention. P300 spellers require multiple iterations for single character selection, and it may take minutes to type a single word[23–29] when not combined with predictive text[30]. Spellers based on the steady-state visual evoked potential (SSVEP), which rely on a combination of gaze shifting and attention-related entrainment of visual cortical neurons to flicker frequency, allow higher information transfer rates (ITRs) due to increased signal-to-noise ratio (SNR) of SSVEPs relative to ERPs[22,31–35]. For instance, Chen et al.[21] and Nakanishi et al.[36] developed SSVEP-based spellers with information transfer rates (ITRs) of ~267 and ~325 bits per minute (bpm) respectively for cued spelling, both with mean classification accuracies of ~90%. These ITRs are the fastest to date, with other state-of-the-art spellers averaging ~146 bpm (108, 124, 144, 151, 167 & 181 bpm respectively)[34,37–41]. This impressive improvement suggests that BCI spellers may become a viable option for hands-free communication outside of clinical settings.

Despite advances in signal processing and classification, BCI spellers have rarely been explored for free communication in healthy individuals. True free communication involves translating momentary thoughts to text, continuously and in real time. However, current tests of "free" spelling often involve users repeating a small number of phrases provided by the experimenter, either from memory or with assistance from salient cues[21,36]. While algorithms may allow ultra-high ITRs on cued tests, it is not clear that naïve users can cope with the significant cognitive load associated with BCI operation to freely communicate at these rates. Consider for instance the virtual keyboard developed by Chen et al.[21], which produced unprecedented ITRs of ~267 bpm. Their approach was to combine joint frequency/phase modulated flicker with filter-bank canonical correlation analysis (CCA), providing high accuracy for large set sizes (40 keys) and short trial durations (1 s). The reported ITRs, however, likely overestimate free communication speed for several reasons. First, the classification assessment consisted of users BCI typing the phrase "HIGH SPEED BCI" three times, interleaved by one-minute breaks. Testing on a single, cued phrase scarcely resembles free communication, which involves the increased cognitive load of generating thoughts, planning phrases, spelling words, and locating the correct characters, all in real-time. Second, ITR was calculated based on a set size of 40 keys. Although all keys were used for template generation, only nine were actually used for typing at test (i.e., "H", "I", "G", "S", "P", "E", "D", "B", "C"), violating the preconditions of ITR calculation[18,42,43]. Finally, in the study of Chen et al.[21], the majority of participants were experienced users, having trained on previous BCI systems, as well as the 200 practice trials in which target letters were highlighted by salient cues. This focus on cued spelling and testing on experienced users may hinder progress toward the development of plug-and-play, brain-based free communication in healthy, naïve users[17,44]. To address this concern, we developed an open-source, high-performance, non-invasive BCI speller and designed a testing protocol to determine its suitability for genuine free communication in naïve users.

A non-invasive interface for brain-based free communication. BCI systems provide a promising avenue for the development of hands-free communication applications for healthy users. We therefore developed a state-of-the-art filter-bank CCA SSVEP BCI speller (Fig. 1)[21,38], examined its suitability for genuine free communication, and evaluated the parameters influencing its performance. In Experiment 1, we tested whether seventeen naïve users could maintain rapid typing during prompted free word association[45]. In Experiment 2, we developed a social BCI communication interface, allowing two users to have a free conversation. To facilitate free communication, we introduced seven important changes to previous top-performing virtual keyboards[21,38].

Reductions in character selection time. (1) We changed the keyboard layout from alphanumeric to QWERTY, which is highly familiar to users[46]. (2) We reduced the number of flicker frequencies (from 40) to 28, excluding numbers and punctuation characters, which were deemed superfluous to free communication. (3) We displayed the three last classified characters at the top of every key, allowing users to monitor decoding while entering virtual keystrokes. This likely reduced working memory load as well as the number of saccades required between classification intervals.

Increases in classification accuracy. (4) To reduce potential interference from endogenous alpha oscillations, we used a higher frequency range (10.0–15.4 Hz rather than 8.0–15.8 Hz). (5) We developed a procedure to calculate the optimal phase shift for the sinusoid templates, thereby tailoring the templates to each individual. (6) We increased the flicker period (from 0.50) to 1.50 s, and the flicker-free interval (from 0.50) to 0.75 s. (7) We developed an averaging method to increase SSVEP SNR, allowing participants with low classification accuracy to potentially use the BCI communication system.

Results
Experiment 1: Results. *QWERTY classification assessment.* Classification templates derived from cued template training ($N = 20$ repetitions/key; Fig. 2a) were evaluated by having naïve participants freely BCI type the complete QWERTY sequence three times, without guiding cues (Fig. 2b). This preliminary test revealed that the BCI naïve participants could use the communication system with varying degrees of voluntary control, with classification accuracy from 22.62–100.00% ($M = 75.37$, $SD = 21.67$), corresponding to ITRs of 9.51–128.2 bpm ($M = 80.41$, $SD = 8.51$; Fig. 3a). All ITRs were calculated based on the method of Wolpaw et al.[18,47], which is commonly used for BCIs[42]. The calculation is as follows:
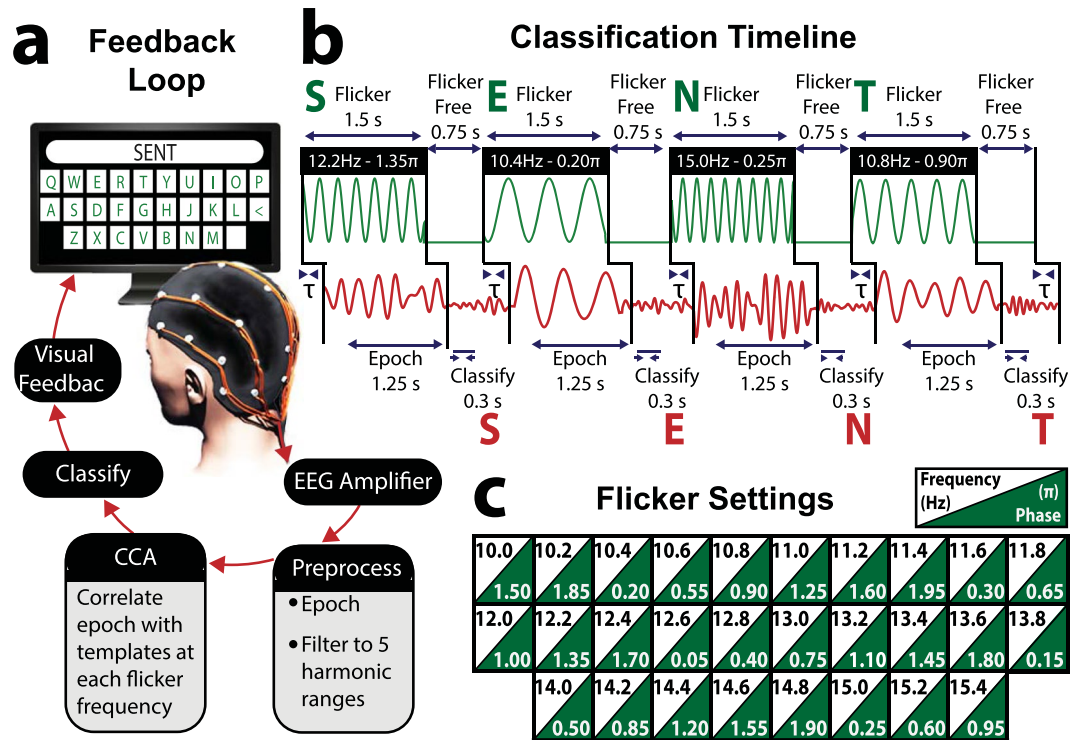
Figure 1. BCI virtual keyboard for free communication. (**a**) Participants operated the real-time feedback loop to freely type words and phrases using their brain activity alone. Participants selected characters in sequence by focusing their attention and fixating their gaze on sinusoidally flickering keys of a virtual QWERTY keyboard on a computer display, which evoked oscillatory SSVEP responses at the corresponding flicker frequency/phase in the EEG. EEG time-locked to flicker was extracted, bandpass filtered to five harmonic ranges and then submitted to a filter-bank CCA with respect to a bank of individualized training templates. The classified frequency was the template most highly correlated with the real-time EEG, with the corresponding character displayed as feedback at the top of each key. Participants were free to select the next character, or to select the backspace key [<] to make a correction. The image of the head was created by Dr. David J. Lloyd. (**b**) Example timeline of visual stimulation and evoked EEG involved in BCI typing of the word "SENT". Each key flickered at a unique frequency/phase for 1.5 s, followed by a 0.75 s flicker-free period, during which the letter was classified and participants shifted their attention to the next key. Focusing attention on a key potentiated the corresponding SSVEP response, increasing the likelihood that the corresponding letter would be classified. τ refers to the SSVEP delay relative to flicker onset, calculated separately for each frequency and harmonic. (**c**) Spatial organisation of the virtual keyboard's flicker frequencies/phases. Each key flickered at a unique frequency/phase, ranging from 10 Hz/1.5π – 15.4 Hz/0.95π.

$$ITR = \frac{log_2N + Plog_2P + (1 - P)log_2\left[\frac{1 - P}{N - 1}\right]}{T} \tag{1}$$

Where N represents the number of possible choices – here $N = 28$ keyboard keys. P represents the classification accuracy, the calculation of which is described for each stage of the experiment in the methods section. Finally, T represents the selection time for each character, which included both the visual stimulation and flicker-free periods for all ITR calculations in this study. Thus, T was 2.25 s (1.50 s stimulation + 0.75 s flicker-free) for both The QWERTY Classification Assessment and BCI Free Communication stages of Experiment 1. Finally, we converted selection times from seconds to minutes to express ITRs in bpm. Reliable free communication was deemed too difficult when classification accuracy was less than 80%. Therefore, only participants with high classification accuracy were considered for the free communication task. Approximately half of the group was classified on either side of this accuracy threshold (accuracy > 80%; $N = 9/17$; $M = 92.29$, $SD = 2.20$ | accuracy < 80%; $N = 8/17$; $M = 57.47$, $SD = 18.13$; Fig. 3a). However, one high accuracy (82.14%) participant elected to undergo retraining rather than free communication, citing frustration with misclassification. Thus, 8 participants performed the free communication task, while the remaining 9 participants underwent retraining instead. The retraining group did not complete free communication, but *instead* undertook a modified template training procedure, described below.

The retraining group ($N = 9/17$) completed template generation based on a double flicker epoch to increase SSVEP SNR. The flicker signal was presented twice in sequence (3.0 s), and the average of the two epochs (1.5 s each) was treated as the single-trial EEG. Again, performance was evaluated by having participants BCI type the
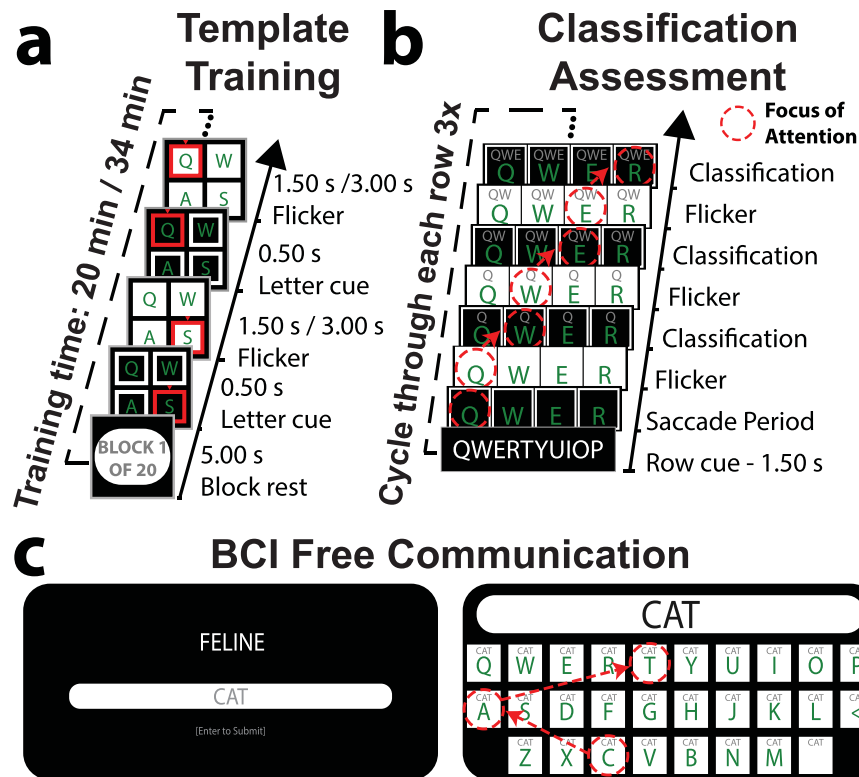
Figure 2. The three phases of Experiment 1, which allowed BCI free communication through prompted free association. (**a**) Template training. Participants ($N = 17$) were cued to focus their attention and gaze on each flickering key in a random order ($N = 20$ repetitions/key). Keys were cued prior to and during flicker. Participants with low classification accuracy ($< 80\%$), determined via QWERTY classification, underwent retraining with an increased flicker duration (3.0 vs. 1.5 s) to improve single-trial SSVEP SNR. (**b**) QWERTY classification. Participants freely BCI typed the complete QWERTY sequence, without guiding cues other than feedback displayed at the top of each key. Flicker-free periods allowed participants 0.75 s to redirect their attention to the next uncued key. The three previous classified letters were displayed at the top of each key, allowing participants to monitor classification while performing keystrokes. (**c**) BCI free communication undertaken by participants with high classification accuracy ($>80\%$). Prompt words allowed participants to freely associate words/phrases. To assess accuracy, participants entered intended character strings using a manual keyboard before BCI typing. A new prompt was presented when participants either matched the intended character string using BCI, or had entered three times more characters than in the intended string.

complete QWERTY sequence three times, without guiding cues (*n.b.* character selection time $T = 3.25$ s [3.00 s stimulation + 0.75 s flicker-free] for the QWERTY Classification Assessment of the re-training group). A paired samples *t*-test demonstrated that classification accuracy (%) was significantly higher for the double ($M = 78.04$, $SD = 19.96$) than single flicker epoch ($M = 60.22$, $SD = 18.85$; $t_8 = -4.34$, $p = 0.002$; Fig. 3b), a mean improvement of ~18%. This increase in classification accuracy resulted in ITRs (bpm) that did not differ significantly for the double ($M = 50.41$, $SD = 6.57$) and single flicker epochs ($M = 54.12$, $SD = 8.41$; $t_8 = 0.75$, $p = 0.476$), despite the much longer flicker duration ($\times 2$). The double flicker epoch resulted in 6/9 participants being classified with greater than 80% accuracy, increasing the system's suitability for free communication. The classification improvement using the double epoch average indicated that filter-bank CCA's algorithms depend on high SNR and single-trial SSVEP phase consistency.

BCI free communication. To determine whether the BCI system was suitable for genuine free communication in naïve users, high accuracy participants ($N = 8/17$) freely generated words/phrases that were semantic associates of word prompts (Fig. 2c; Supplementary Information, Appendix 1). Participants first manually typed their intended words/phrases using a standard keyboard, and then attempted to replicate these character strings using BCI typing. This allowed us to quantify classification accuracy during free communication.

Participants could successfully freely communicate using the BCI system, generating a large variety of unique words and phrases in response to the prompts (Supplementary Information, Appendix 2). Example prompt/associate pairs include: "GREAT"/"DIM SUM", "IMPORTANT"/"MUM" and "PLACE"/"IT PUTS THE LOTION ON ITS SKIN". Participants generated 1–10 words/prompt ($M = 1.80$, $SD = 0.53$), selecting each key at least once ([X] & [Z]) and up to 201 times ([backspace]; $M = 21.26$ selections, $SD = 8.77$). Participants generated words of average complexity (word length: $M = 5.13$ characters, $SD = 0.22$), equivalent to the average length of English words (5.10 characters)[48]. Free communication was perhaps most strongly evident in the ability of many prompts (selected at random) to produce different successfully BCI typed associates across participants (e.g., prompt:
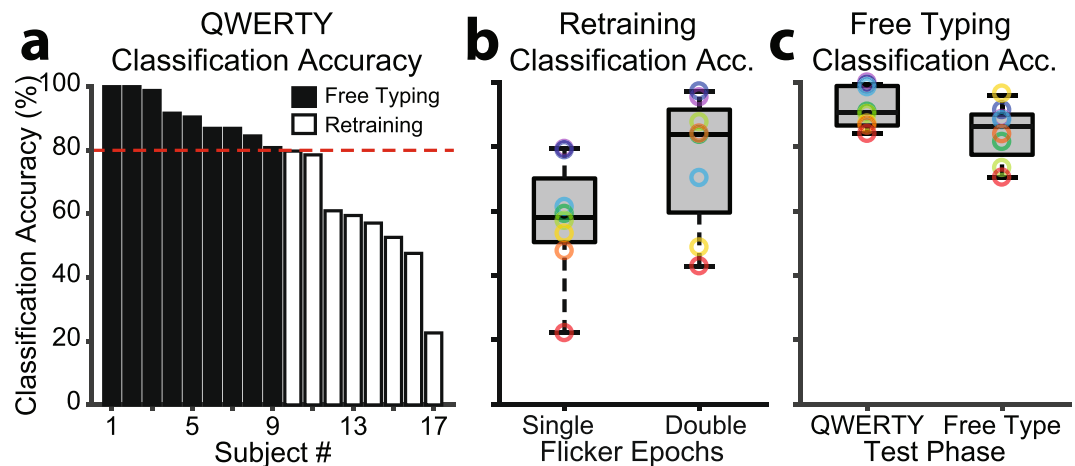
Figure 3. Accuracy during QWERTY classification and free communication. (**a**) Individual differences in QWERTY classification (single flicker epochs: 1.5 s). Participants classified with greater than 80% accuracy ($N = 8/17$) advanced to free communication, with the remaining participants undergoing retraining with a double flicker epoch (3 s). (**b**) Box plot of QWERTY classification for the low accuracy group with 1.5 s single and 3 s double flicker epochs. (**c**) Box plot of performance for the high accuracy group on QWERTY classification and free communication via prompted association. In these and subsequent box plots, coloured rings represent individual participants, the central bars represent the median, outer bars represent the 25th and 75th percentiles, and the whiskers extend to the absolute maxima.

"END"; associates: "OF THE DAY", "FINAL", "HOLIDAY", "START"). This indicated that communication depended on the participants' individual real-time thoughts and that the association task successfully tapped free communication.

Classification accuracy and ITR are often determined by instructing users to repeat phrases or cycle systematically through the keyboard. We included such a classification assessment, with the aim of contrasting free communication performance. ITRs during free spelling (92.41 bpm) were lower than during this systematic QWERTY assessment (109.56 bpm). Indeed, a paired samples t-test revealed that classification accuracy (%) was significantly lower during free communication ($M = 84.22$, $SD = 3.09$) than on the QWERTY assessment ($M = 92.29$, $SD = 2.20$; $t_7 = 2.96$, $p = 0.021$; see Fig. 3c). The performance cost of free communication (~8%) might be partly attributable to increased demands on memory and search during free communication, and indicates that instructed assessments overestimate free communication ITRs.

Factors affecting the feasibility of BCI free communication. As classification accuracy varied substantially across participants, we investigated a number of contributing factors. To evaluate SNR, we examined average FFT amplitude spectra for each frequency during the (single epoch) template generation phase, undertaken by both the low and high accuracy groups. Averaged SSVEPs for each frequency ($N = 20 \times 1.5$ s epochs) were zero-padded to 5.0 s to allow 0.2 Hz spectral resolution and separation of adjacent flicker frequencies (Fig. 4a). To evaluate the effect of SNR on classification accuracy, we calculated the difference in FFT amplitude spectra between the high and low accuracy groups (Fig. 4b). SSVEP amplitudes were larger for the high compared with low accuracy group, especially for the first three harmonics. The grand mean SSVEP amplitude topographies of the first harmonics revealed maximal amplitudes at occipitoparietal sites, consistent with previous frequency tagging studies of attention[49–51], with larger amplitudes for the high compared with low accuracy group (Fig. 4c). This effect was confirmed statistically: classification accuracy (%) was strongly positively correlated at the between-subjects level with the mean SNR of the first harmonic flicker frequencies ($r_{15} = 0.84$, $p < 0.001$; Fig. 4d). These results demonstrate that SSVEP SNR is critical for reliable free communication.

Related to SSVEP SNR, the range of stimulation frequencies is a free parameter likely to affect classification accuracy. Assessment of the first harmonic SSVEPs revealed that SNR was lower for frequencies in the alpha range (10.0–12.0 Hz) than for higher frequencies (12.0–15.4 Hz; Fig. 4e). This may owe to interference at lower frequencies from phase-misaligned endogenous alpha oscillations[52]. Consistent with this interpretation, across the cued frequency spectrum (10–15.8 Hz), classification features (R_{fi}) on the QWERTY assessment were numerically higher for template frequencies in the alpha range (10–12 Hz) compared with those above (>12 Hz). Note that these classification features reflect the statistical similarity between the real-time EEG and the training templates of each frequency (see also Experiment 1: Methods; Individualized Filter-Bank CCA Classification). Thus, these patterns of higher classification features in the alpha range reflect an increased likelihood of the real-time EEG being misclassified as an alpha frequency. The effect of low SNR for classification features in the alpha range was apparent only for the low accuracy group (Fig. 5a), increasing the likelihood of misclassification specifically for low accuracy individuals. Thus, avoiding stimulation frequencies in the alpha range might increase the feasibility of free communication.

Our filter-bank CCA algorithms used multiple stimulation frequency harmonics. *Prima facie*, the inclusion of additional harmonics might improve accuracy, as classification is based on more information. To examine
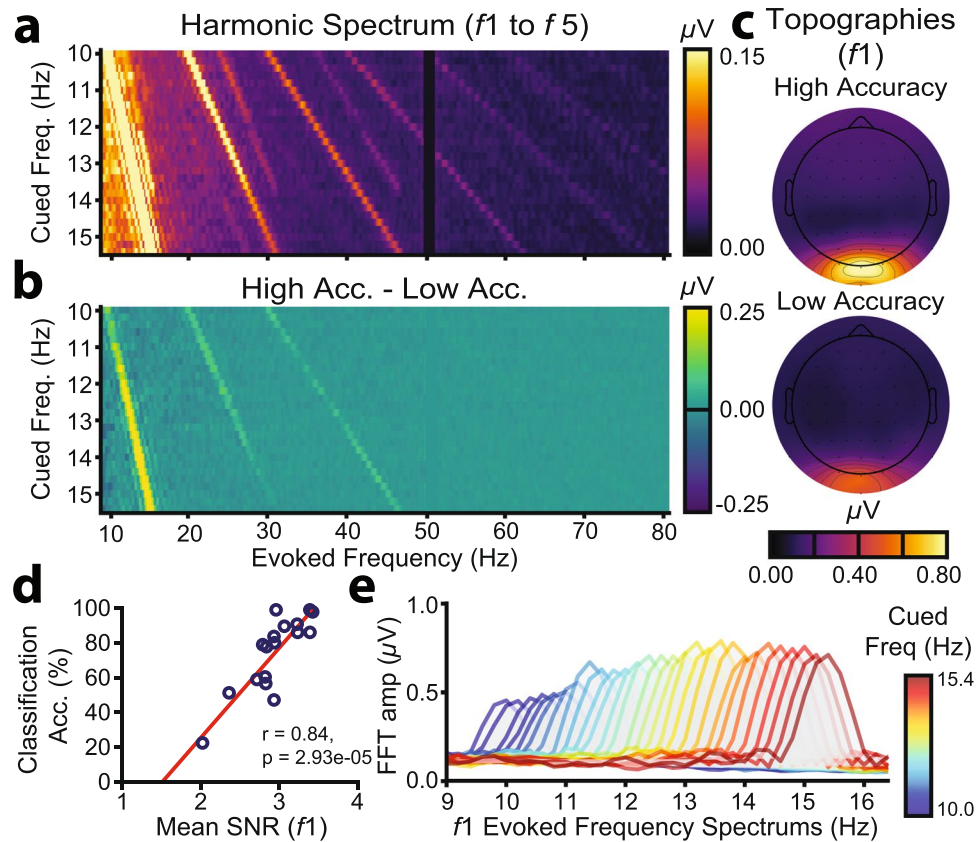
Figure 4. SSVEP SNR. (**a**) Grand mean FFT ERP amplitude spectra across all participants plotted for harmonics 1–5. Warmer colours indicate higher SSVEP amplitudes. The colour map is scaled to highlight later harmonics; the slice through the first harmonic in fact extends to 2.10 μV (**b**) Differences in grand mean spectra for the high and low accuracy groups. (**c**) Grand mean SSVEP amplitude topographies at the first harmonic, averaged across all flicker frequencies, plotted separately for the high and low classification accuracy groups. (**d**) Scatter plot of the positive relationship between mean SSVEP SNR at the first harmonics and QWERTY classification accuracy. (**e**) Grand mean FFT amplitude spectra at the first harmonic for each cued frequency during template training.

this, we classified the baseline template data ten times, incrementally including an additional harmonic at each step (i.e., [1]: f_1; [2]: f_1, f_2; [3]: f_1, f_2, f_3 …). As depicted in Fig. 5c, classification accuracy in fact benefited from fewer rather than more harmonics. To assess this statistically, classification accuracy (%) was submitted to a two-way mixed ANOVA with group (high accuracy, low accuracy) and number of harmonics (1, 2, 3, 4, 5, 6, 7, 8, 9, 10) as factors. There were significant main effects of group ($F_{1,15} = 25.95$, $p < 0.001$, $\eta_p^2 = 0.63$) and number of harmonics ($F_{9,21} = 8.05$, $p = 0.005$, $\eta_p^2 = 0.35$; Greenhouse-Geisser adjusted, Mauchly's $W < 0.001$, $X^2 = 263.23$, $p < 0.001$), but no significant group × number of harmonics interaction ($F_{9,21} = 0.39$, $p = 0.388$, $\eta_p^2 = 0.06$). Thus, the effect of the number of harmonics was statistically similar for the high and low accuracy groups. Importantly, the main effect of harmonic was better explained by a quadratic ($F_{1,15} = 37.60$, $p < 0.001$, $\eta_p^2 = 0.638$) than linear trend ($F_{1,15} = 4.08$, $p = 0.062$, $\eta_p^2 = 0.214$), confirming an overall benefit for fewer harmonics. Examination of the individual-level data indicated that classification accuracy peaked for some participants after only one harmonic, while others benefited from the inclusion of up to seven harmonics (see Table 1 for the optimal number of harmonics for each participant). Classification accuracy was significantly higher when using each individual's optimal number of harmonics ($M = 79.55$, $SD = 20.83$) than the five harmonics used in real-time ($M = 75.35$, $SD = 21.69$; $t_{16} = -3.36$, $p = 0.004$; Fig. 5b), although modestly so (~4%). Thus, free communication can be facilitated by selecting the optimal number of harmonics separately for each individual.

Joint frequency/phase modulated flicker with filter-bank CCA allows high ITRs but can require lengthy individualized calibration. In Experiment 1, template generation required ~18.67 minutes of visual stimulation (2 s/trial × 28 frequencies × 20 trials/frequency). To determine whether shorter calibration is possible, we retrained the filter-bank CCA 20 times for each participant, incrementally including an additional trial for each classification. Using MATLAB's *fittype* and *fit* functions, which apply the method of least squares, we fitted an inverse exponential function to the resulting classification accuracy curve (Fig. 5d):
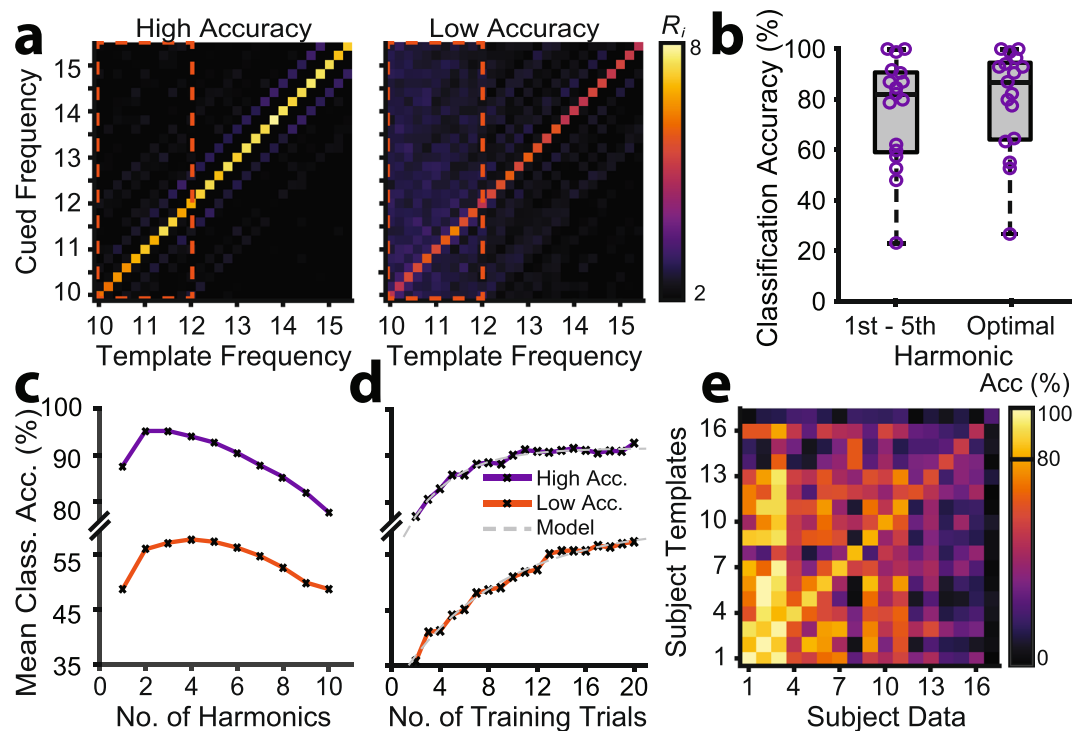
$$ACC = a(1 - e^{-s(T-i)}) \qquad (2)$$

Figure 5. Factors affecting the feasibility of BCI free communication, assessed using QWERTY classification data. (**a**) Grand mean classification features (R_i) for each cued frequency for the high and low accuracy groups. For each flicker period, the filter-bank CCA produced a classification feature (R_i) representing the correlation between the single-trial EEG and templates at each frequency. Correct classification occurred when R_i was maximal for the template matching the cued frequency. Template frequencies in the alpha range are indicated by the orange bounding boxes (dashed lines). (**b**) Box plots showing the improvement in classification accuracy using the optimal number of harmonics rather than the fixed first five harmonics that were used in real-time. (**c**) Simulated classification accuracy by number of harmonics, plotted for the high and low accuracy groups. (**d**) Simulated classification accuracy by number of training trials, plotted for the high and low accuracy groups. The fitted model is an inverse exponential function. (**e**) Cross-participant classification. To assess template generalisability, each participant's single-trial EEG was classified using all other participants' templates. The leading diagonal represents accuracy when participants were classified using their own templates. The leftmost column (read bottom to top) represents the highest accuracy data classified with progressively lower accuracy templates. Similarly, the bottom row (read left to right) represents the highest accuracy templates used to classify progressively lower accuracy data.

Where *ACC* is classification accuracy, *T* is the number of training trials, *a* is the asymptote, *s* is the scaling factor and *i* is the *x*-axis intercept[53]. On average, the high accuracy group required 12 training trials to reach 99% of their asymptote (90.34% accuracy). In contrast the low accuracy group was projected to require 33 training trials to reach 99% of their asymptote (60.21% accuracy). This suggests that high SNR users require fewer template generation trials for reliable free communication, and that template generation duration can be decreased for these individuals.

While reducing training is desirable, it would be ideal to eliminate training entirely. We therefore assessed whether participants could be cross-classified with other participants' templates. If cross-classification is feasible, the BCI communication system could be used by new users without individualized calibration. As is evident from the results of the cross-classification analyses (Fig. 5e), cross-classification depends on the to-be-classified individual's own accuracy. Specifically, participants with high classification accuracy were accurately classified with other participants' templates, while participants with low classification accuracy could not be classified well, even with high accuracy templates. For example, close inspection of Fig. 5e reveals that participant #3 could be classified at over 80% accuracy using 13/17 participants' templates. There were three cases in which classification accuracy was as good or better using another person's templates compared with one's own. This suggests that high SSVEP SNR participants could potentially forego template generation entirely and instead freely communicate from the outset using generic templates.

Experiment 2: Results. BCI communication systems should be useful not only for expressing thoughts, but also for exchanging thoughts explicitly in conversation with others. To fully evaluate the system's efficacy for free communication, we therefore extended the system to include an asynchronous two-user messaging inter-face (Fig. 6). Using the default parameters of Experiment 1 (namely, 5 harmonics & 20 training trials for each BCI key), we calculated classification accuracy on the template training data by excluding the to-be-classified EEG from template SSVEP generation. Due to computational practicalities, the optimal template sinusoid phases

Participant	Real-time (%)	Accuracy using Optimal Number of Harmonics (%)	ITR using Optimal Number of Harmonics (bmp)	Optimal Number of Harmonics (N)
P06	100.00	100.00	128.20	5
P12	100.00	100.00	128.20	1
P02	98.81	98.81	124.20	2
P08	91.67	92.86	109.24	2
P16	90.48	96.43	117.74	3
P05	86.90	90.48	104.02	2
P13	86.90	94.05	111.97	1
P09	84.52	86.90	96.66	3
P17	82.14	92.86	109.24	2
P10	79.76	82.14	87.50	7
P04	78.57	79.76	83.16	4
P01	61.90	63.10	56.07	6
P07	59.52	64.29	57.84	1
P03	57.14	77.38	78.95	2
P11	52.38	52.38	41.19	5
P15	47.62	54.76	44.34	2
P14	22.62	26.19	12.49	6

Table 1. Real-time and offline optimized classification accuracy (%) on the QWERTY classification assessment.

were calculated using all the training data (see also Experiment 1: Methods; Template Training). Using this procedure, classification accuracy on the template training data was 96% (ITR = 131.25 bpm; excluding reading time) for participant one (P1) and 98% (ITR = 137.12 bpm; excluding reading time) for participant two (P2), suggesting that the classification algorithms were ready to be applied in a hands-free conversation (~5.7 words/minute). Note that these baseline/offline ITRs were calculated using a character selection time T of 2.00 s (1.50 s stimulation + 0.50 s flicker-free). The two participants freely conversed using the BCI messaging interface for ~55 minutes, using an EMG enter key [↵] activated by jaw clenching to complete their messages, view the messaging display and recommence BCI typing. We used muscle rather than brain activity for the submit key to keep the number of BCI keys and corresponding CCA classes ($N = 28$) consistent between Experiments 1 and 2, and to ensure that chat text was terminated with maximum reliability/control. The high reliability of the enter key allowed us to calculate BCI classification accuracy using backspace keystrokes as a proxy for classification/user errors. For consistency with Experiment 1 and previous reports, ITRs were calculated with respect to the BCI keys and did not include EMG enter keystrokes. In developing one of the first BCI spellers for free conversation, we found that a hybrid EEG/EMG approach was more robust and usable than pure EEG. The participants' conversation focused on recent and upcoming social engagements and their immediate experience using the social interface (see Supplementary Information, Appendix 3 for an unedited transcript). In total the two participants generated 68 messages (P1: 33; P2: 35), including 349 words (P1: 170, P2: 179), 1,731 characters (P1: 824, P2: 907) and 283 spaces (P1: 139, P2: 144). Mean word length was 4.2 characters ($SD = 2.2$, range = 1–13), and was similar for the two participants (P1: $M = 4.0$, $SD = 2.1$, range = 1–13; P2: $M = 4.3$, $SD = 2.3$, range = 1–13). To evaluate the conversation's trajectory during recording, the topic of each message was qualitatively scored (Fig. 7a). Nine conversation topics were identified, with each participant contributing at least one message to each topic. Interestingly, the topics overlapped in time: at least two and as many as five topics were ongoing concurrently. This appeared to reflect an emergent feature of the interface, allowing each participant to BCI type continuously, rather than iteratively awaiting their partner's reply.

For both participants, there was a statistically reliable positive Pearson correlation between message length and ordinal message number (P1: $r_{31} = 0.39$, $p = 0.027$; P2: $r_{33} = 0.50$, $p = 0.002$; Fig. 7b), suggesting that the participants gained confidence in the interface over the course of recording. The participants on average spent 78% of the recording duration BCI typing, with the remaining 22% devoted to viewing typed messages. The considerable proportion of time spent viewing messages indicated that ITRs calculated offline overestimate pure character transfer during free conversation, which naturally entails turn taking.

As depicted in Fig. 7c, the backspace and space keys were selected most frequently by both participants, and together the participants used each BCI key at least once. To test whether the two participants differentially selected particular keys, chi-square tests of independence set expected values (N) at half the total number of observations for each key. Twenty-four of the 29 keys (including the EMG enter key [↵]) had sufficiently large expected values ($N > 5$) to test statistically, excluding keys [Q], [J], [Z], [X] and [V], which were selected too infrequently. Observed and expected values did not differ significantly for any of the keys (χ^2s < 3.90, ps > 0.0484; $\alpha = 0.002$; Bonferonni $\alpha = 0.05/29 = 0.002$), indicating that key selection was similar for the two participants during free communication. Counting backspace corrections as classification or participant errors, overall classification accuracy during free communication was 88% (ITR = 98.86 bpm) for P1 and 90% (ITR = 103.01 bpm) for P2, down 8% from training classification. Online ITRs for free communication were calculated using a character selection time T of 2.25 s (1.50 s stimulation + 0.75 s flicker-free). These results show that the BCI messaging
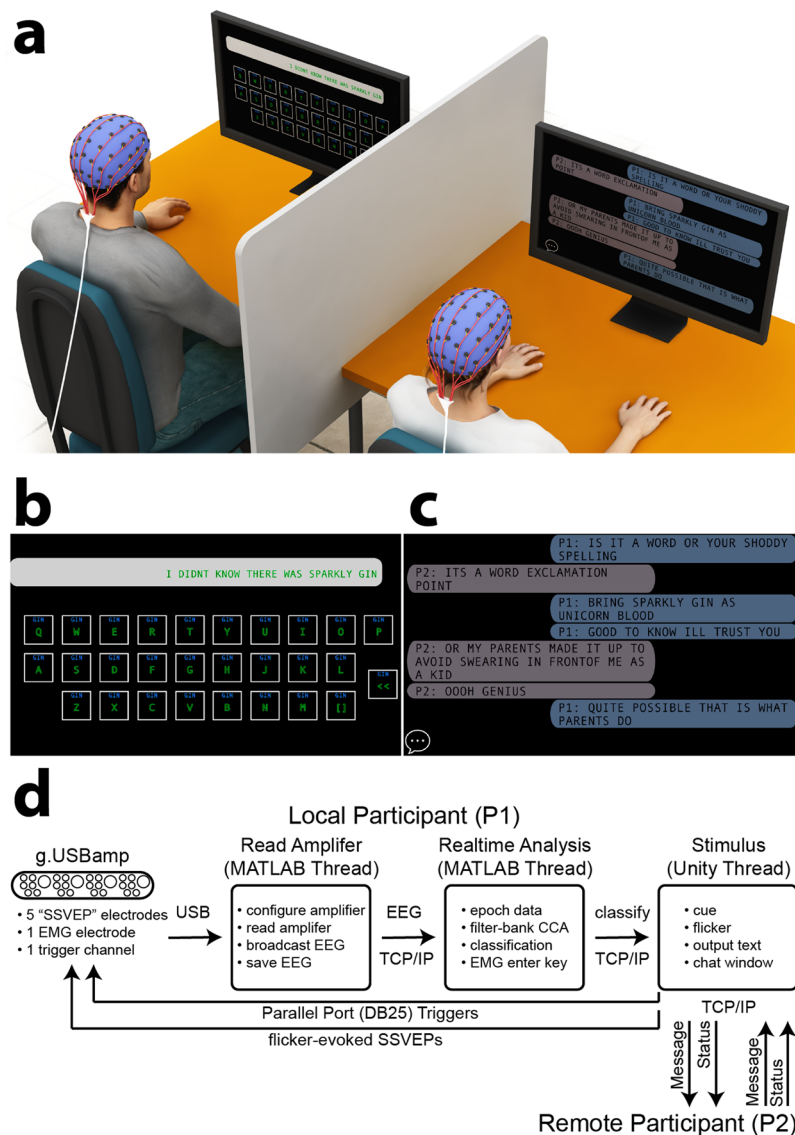
Figure 6. Experiment 2: BCI communication system for free conversation. (**a**) Two experienced participants used an asynchronous messaging interface to have an unprompted free conversation using a system with only six electrodes. This image was created by Dr David J. Lloyd. (**b**) The keyboard layout was similar to Experiment 1. The interface additionally included an electromyography (EMG) enter key [↵], controlled by detecting jaw clench signals at a frontal scalp electrode, allowing the participants to complete their messages, view the messaging display and recommence BCI typing. (**c**) BCI messaging display. The local participant's (P1′s) messages are indicated with light blue, and the remote participant's (P2′s) with light grey. The chat icon is enabled (bottom left corner), indicating that P2 is currently BCI typing, rather than viewing messages. (**d**) Three parallel threads devoted to reading the EEG, real-time analysis and stimulus presentation (indicated by rounded boxes). Inter-thread communication was via parallel port and TCP/IP (indicated by arrows). "Message" indicated the BCI typed text, and "status" indicated whether the remote participant was typing or viewing messages. The flow of information between threads was identical for P2.

interface was suitable for free communication, and that offline classification performance reflects an upper estimate for ITR during free communication.

Discussion

Free and open communication is central to modern civilisation, allowing people to convey their thoughts and interface with computers and the internet. While individuals are remarkably adept at operating manual keyboards, a next frontier is communication without manual input. Here, we developed a high performance SSVEP speller based on filter-bank CCA, examined its feasibility for free communication, and evaluated the free parameters that could be altered to optimise performance. In Experiment 1, we tested whether naïve users could maintain rapid typing during prompted free word association. In Experiment 2, we developed a social messaging interface, allowing two users to have an unprompted free conversation. Overall, our results showed that traditional cued
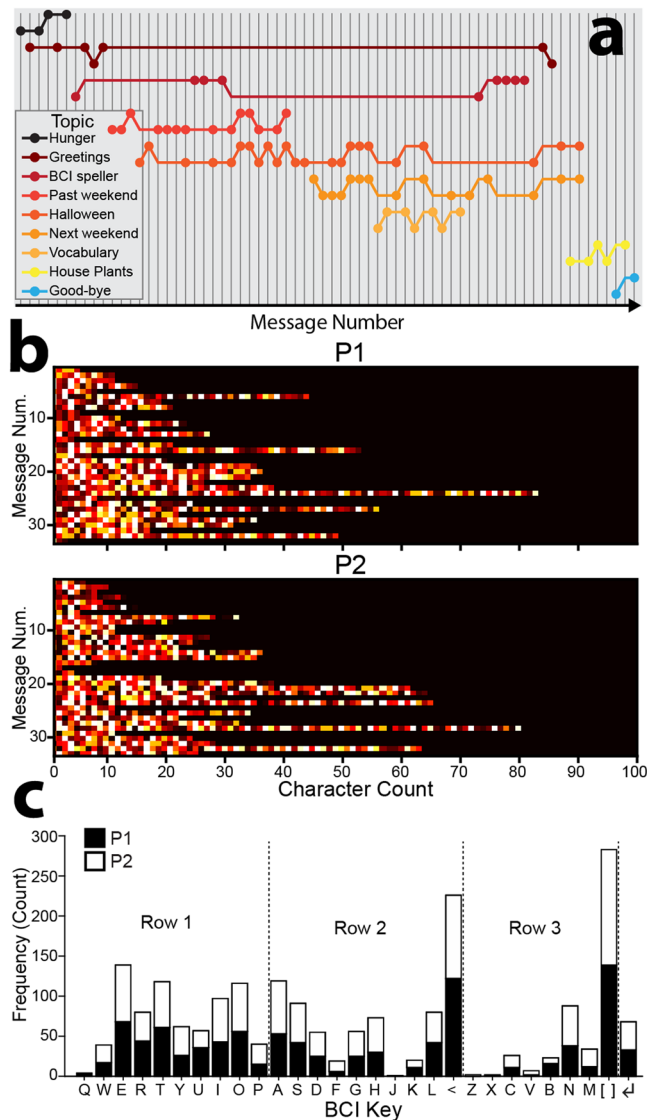
Figure 7. Free communication results of Experiment 2. (**a**) Streams of brain-based free conversation. Each message was qualitatively scored as belonging to one of nine conversation topics. Each colour represents a topic, and each vertical grey line represents a message. Circles higher in the conversation thread for each topic represent messages sent by P1, while lower circles represent messages sent by P2. (**b**) Messages sent for the two participants (P1 & P2) as a function of ordinal message number and character count. Brighter colours indicate selection of keys later in the keyboard (row three). (**c**) Key selection counts for the two participants, including an EMG enter key [↵] activated by jaw clenching.

typing tests overestimate free communication ITRs (Fig. 3c). However, given individualised interfaces involving sufficient template training trials (Fig. 5d) and flicker durations (Fig. 3b) and appropriately chosen harmonic (Fig. 5b,c) and frequency (Figs. 4e and 5a) parameters, the majority of naïve users would be able to freely communicate. The single greatest determinant of free communication success was SSVEP SNR, which showed a strong positive correlation with classification accuracy (Fig. 4d). Our results suggest that individuals with high SNR might begin free communication earlier in recording due to the reduction of training trials and the ability to be cross-classified with generic templates (Fig. 5e). The successful brain-based free conversation of Experiment 2 (Supplementary Information, Appendix 3) suggested that experienced users with high SSVEP SNR could use the virtual keyboard as naturally as a manual keyboard, albeit more slowly. Message character counts increased for both participants during their free conversation (Fig. 7b), suggesting that experienced users can improve their BCI communication efficiency, even when classification parameters remain constant.

Our communication system was made possible by the strong foundations and remarkable recent progress in the field of non-invasive BCI spellers. The first BCI speller, based on the P300 ERP, could reliably classify characters at a rate of 12 bpm (~2.3 characters per min)[25]. Modern spellers using SSVEPs, which offer higher single-trial SNR relative to classical ERPs, can achieve rates of around 146 bpm, an order of magnitude faster than early spellers[34,37–41]. The seminal development of filter-bank CCA has allowed the report of an unprecedented ITR of 267

bpm, reflecting a forward step in non-invasive neuroimaging[21,38]. However, these impressive leaps in ITR were predominantly calculated using generic, cued character strings. Ultimately, BCI systems are intended for free communication, which is inherently interactive and spontaneous.

Our studies investigated the usability of BCI systems for free communication. As a first step, we developed three tests with increasing levels of user freedom and expression: (1) QWERTY classification in which users were instructed but not explicitly cued to complete the entire keyboard sequence. In conjunction, we introduced status characters at the top of each key that allowed participants to track classification while entering keystrokes. (2) We developed a prompted free association task in which the ground truth for character intention was established by having users enter the target phrase using a manual keyboard before BCI typing. Free association allowed users to generate their own words and phrases with minimal external input. (3) We introduced free brain-based communication, allowing users to converse freely, without input from the experimenter. To support free conversation, we developed a social BCI interface that allowed users to view and respond to their conversational partner's messages, received asynchronously. Our results support the use of joint frequency-phase modulated CCA BCI systems for free communication, but indicate that ultra-high ITRs are not realistic for free communication given current interfaces, which require serial key selection and lack predictive text.

Our results indicate that effective free communication requires a focus on usability rather than fast character selection time. Pilot development indicated that naïve users could not reliably make saccades to the next key with the short trials durations (i.e., 0.5 s flicker/0.5 saccade) employed in previous work[21,38], especially during free communication. For these users, short durations were not conducive to free communication due to the overwhelming cognitive load of focusing selective attention on a virtual key, ignoring distraction from adjacent flickering keys, planning successive keystrokes, locating the next key and making backspace corrections for misclassifications. Even with longer character selection times (1.5 s stimulation/0.75 s flicker-free), the effect of the added cognitive load during free communication is illustrated by the ~8% reduction in classification accuracy experienced by the eight naïve users of Experiment 1 who progressed from the QWERTY assessment to free communication. Consistent with this result, the two experienced users in Experiment 2 also showed an ~8% reduction in classification accuracy from offline assessment to free communication. This suggests that cued and instructed typing tests overestimate free communication ITRs. As an alternative explanation, the apparent classification accuracy cost for free communication might instead reflect an overestimation of baseline classification accuracy[54,55]. Note though that our primary goal was to test whether free communication between healthy users using SSVEP BCI might be possible with relatively high ITRs. Having achieved this aim, our work paves the road for future studies investigating the brain activity patterns unique to free communication compared with cued spelling[17]. We believe that the observed reductions in classification accuracy during free communication may have been more drastic if not for measures such as reducing the number of keys, using a QWERTY layout and displaying classification feedback on each key.

Improving the usability of the BCI speller as a hands-free communication device involves optimising classification accuracy. Individual differences in classification accuracy were largely attributable to SSVEP SNR. Participants with low SNR were retrained using a double flicker epoch, with classification based on the mean SSVEP of the two epochs. This nearly doubled character selection time, but greatly increased classification accuracy (+18%), allowing ITRs to remain constant. Thus, a focus on usability over character selection time provides ITRs sufficient and practical for free communication. Indeed, as users anecdotally remarked through BCI typing: "I WANT ONE OF THESE ON MY PHONE" and "TYPING WAS NEVER BETTER". Ultimately, usable interfaces for free communication require serviceable rather than ultra-high ITRs.

A main advantage of communication systems based on filter-bank CCA is that the analysis parameters can be adapted and individualised to optimise performance. Our results showed that the most reliable determinant of classification accuracy was SSVEP SNR. A relatively simple procedure for optimising performance would be to determine SNR across a range of frequencies before template training[56]. This could help select the optimal frequency range, which would be especially advantageous for participants with low SNR in the alpha range. Determining the SSVEP SNR early could allow high SNR participants to proceed immediately to free communication using generic templates, while low SNR participants might undergo double epoch training. Additionally, template training could be optimised by real-time modelling of increases in classification accuracy with additional training trials, which we show is well-characterized by an exponential function that approaches an asymptote. Therefore, real-time evaluation using principled stopping rules might optimise accuracy and minimize training time. Real-time evaluation might also determine the optimal number of harmonics for each individual. Together, these additional measures would optimise performance and reduce training time, improving usability and end-user experience.

The rise of virtual/augmented reality has created new opportunities for BCI communication systems beyond their clinical origins. For instance, future systems might allow everyday users to communicate with peers as they navigate virtual worlds, or allow others to discretely compose emails while walking to their next business meeting. The development of such real-world BCI systems requires usability-centred design. For example, in developing BCI for free conversation in Experiment 2, we found that the system was more robust/usable using hybrid EEG/EMG rather than pure EEG. Here, BCI spelling was based on EEG activity alone, and the two users submitted their entries using an EMG enter key, based on jaw muscle activity. Using EMG ensured that chat text was terminated with maximum reliability/control. Our communication system functions as an early prototype for general-purpose use in naïve users, with many open paths. Development could focus, as we have, on non-invasive sparse electrode systems, which are suited to affordability and portability. Our results demonstrate that well-designed sparse electrode systems can provide high-performance. Real-time classification accuracy was ~89% for free conversation using only five classification electrodes. We show that individualizing the system can improve classification performance by as much as 18%. Further, adaptive interfaces might maximize efficiency using general-purpose templates, as shown by our cross-classification analysis, or introduce new features such as

Description	URL
Experiment 1/2 Source	https://github.com/davidrosspainter/brain-to-brain-interface/
Experiment 1 Data	https://osf.io/pu3v9/files/
Experiment 2 Data	https://osf.io/xsq8y/files/

Table 2. Online source code and data repositories. Source code includes all template training, real-time classification and visual stimulation routines.

predictive text[30,57–61] or mental imagery decoding[62] to narrow the search space of possible user intentions, increasing efficiency especially for low SNR users.

BCI communication systems have applications hitherto futuristic, made possible by recent advances in signal processing and decoding of neural activity patterns. We have shown that appropriate improvements to existing systems allow major increases in usability, here enabling free communication in naïve users. More specifically, given individually tailored analysis parameters and explicit usability design, filter-bank CCA provides a powerful basis for robust BCI free communication. To explore this possibility further, we recommend that performance appraisals of future systems reflect their intended modes of naturalistic free communication and control.

Methods

Experiment 1: Methods. *Participants.* Seventeen BCI naïve participants (7 males, age $M = 25.12$ years, $SD = 6.82$) volunteered after providing informed written consent and were paid \$40. All participants were highly familiar with the QWERTY keyboard layout (typing speed $M = 231.76$ characters per min, $SD = 52.62$ characters per min; ITR = 1152 bpm) but naïve to the speller. The mean typing speed for the group ranks at the 59th percentile of over 46 million runs (https://typing-speed-test.aoeu.eu/). Experiments 1 and 2 were approved by The University of Queensland Human Research Ethics Committee and were performed in accordance with the relevant guidelines and regulations. The participants provided informed written consent to have their deidentified data made open access (Table 2).

Overview of BCI communication system. The system implemented joint frequency/phase modulated flicker paired with a filter-bank CCA (Fig. 1a,b). The joint frequency/phase modulated flicker method relies on constant electrophysiological latency across stimulation frequencies, and sets similar flicker frequencies at uncorrelated phases. Filter-bank CCA improves the classification accuracy of standard CCA by using the SSVEP harmonics in combination with the fundamental frequencies. In this study, 28 virtual keys ([A]–[Z], [SPACE] & [BACKSPACE]) were arranged in a QWERTY keyboard layout. Each key was tagged with sinusoidal flicker at a unique frequency/phase (Fig. 1c).

Experiment protocol. **Phase 1: Template Training**. This procedure was used to generate individualised templates of the neural activity evoked by focusing on each key in the virtual keyboard. A red outline and arrow cued participants to foveate and focus their attention on each key. Each key was cued 20 times in a random order. Keys flickered for 1.5 s followed by a 0.5 s flicker-free period, during which eye movements could be made to the next letter (see Fig. 2a). A 5 s rest period followed each cycle through the keys. The recording duration was ~20 minutes, including rests.

Phase 2: QWERTY Classification Assessment. This phase determined classification accuracy using the training templates. Participants focused on each key for one flicker period, cycling through each row from left to right/ top to bottom, starting at [Q] and ending at [SPACE]. Participants cycled through the keyboard three times. Importantly, typing was self-directed as no cue was presented to direct participants' attention to the correct key. Instead, feedback was provided by status characters reflecting the last three classifications, printed at the top of each key (Fig. 2b). Classification accuracy (P) was calculated by dividing the number of epochs during which the correct/expected letter in the sequence was classified by the total number of epochs in the sequence. The flicker-free period was set to 0.75 s. Letter classification took ~0.30 s to compute; thus participants had an additional 0.45 s to redirect their gaze if necessary, which was ample time to complete an eye movement[63]. If classification accuracy in the testing phase was greater than 80%, participants proceeded to free communication (phase 3). Otherwise, participants were deemed unable to reliably communicate and underwent retraining in which two 1.5 s flicker periods were concatenated, with a phase reset after the first 1.5 s. The two 1.5 s epochs were averaged together, which improved the SNR for these individuals.

A challenge for free communication is locating and fixating on the correct key before the onset of the next flicker epoch. Participants were therefore given three minutes to practice using the BCI system before the free communication task began.

Phase 3: BCI Free Communication. This phase assessed the BCI system's suitability for free communication. Participants were instructed to generate responses to prompt words in a free association task[45]. At the beginning of each trial, participants were presented with a prompt word, which was randomly selected from a list of 321 common English words (Supplementary Information, Appendix 1; selected from: https://www.ef-australia.com.au/english-resources/english-vocabulary/top-3000-words/). Participants used a physical QWERTY keyboard to manually type the first word or phrase which came to mind upon seeing the prompt. Participants then attempted to replicate the character string using the BCI system (Fig. 2c). When the BCI typed character string matched that submitted using the manual keyboard, the trial ended and a new prompt was presented. If participants were unable to replicate the target character string after entering more than three times the target number of characters, the trial was aborted and a new prompt was presented. Participants performed the free association task
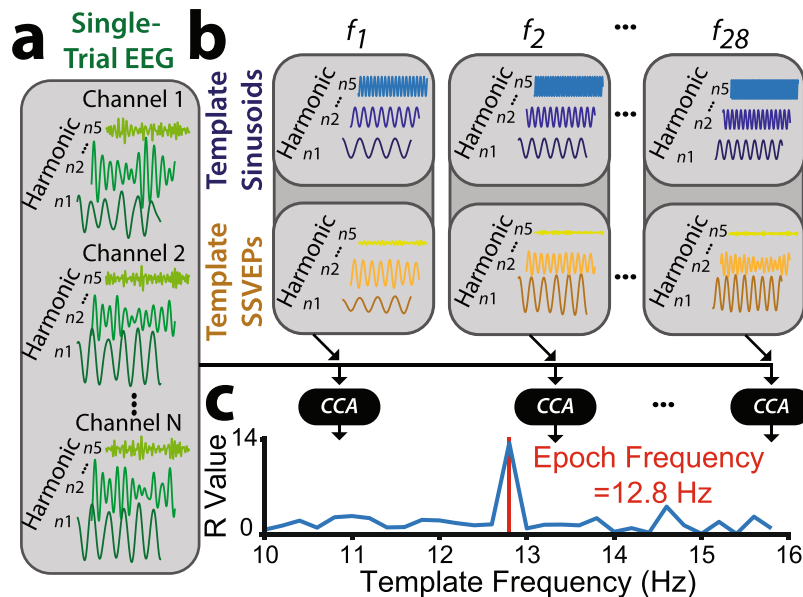
Figure 8. Individualised filter-bank CCA classification. (**a**) Following each flicker period, the corresponding single-trial EEG from the occipitoparietal electrodes with the highest SSVEP training amplitudes were bandpass filtered to the first five harmonic ranges. (**b**) Training resulted in template SSVEPs and sinusoids, reflecting mean signals for each of the 28 flicker frequencies and first five harmonics (see Methods: Template Training for a complete description). (**c**) Filter-bank CCA evaluated the overall pattern of five weighted correlations between single-trial EEG, template sinusoids and template SSVEPs. The filter bank CCA produced a final outcome feature for each frequency (R_{f_i}), representing the degree of similarity between the single-trial EEG and templates. The frequency with the highest similarity was classified as the selected frequency/key.

for 30 minutes. Classification accuracy (*P*) was calculated by comparing BCI-entered characters with the target string. Superfluous characters were counted as errors.

EEG Recording and Channel Selection. EEG data were sampled at 2048 Hz using a BioSemi Active Two amplifier (BioSemi, Amsterdam, Netherlands) from 64 active Ag/AgCl scalp electrodes arranged according to the international standard 10–20 system for electrode placement in a nylon head cap[64]. The electrode positions were: AF3, AF4, AF7, AF8, AFz, C1, C2, C3, C4, C5, C6, CP1, CP2, CP3, CP4, CP5, CP6, CPz, Cz, F1, F2, F3, F4, F5, F6, F7, F8, FC1, FC2, FC3, FC4, FC5, FC6, FCz, FP1, FP2, FPz, FT7, FT8, Fz, Iz, O1, O2, Oz, P1, P10, P2, P3, P4, P5, P6, P7, P8, P9, PO3, PO4, PO7, PO8, POz, Pz, T7, T8, TP7 and TP8. The common mode sense (CMS) active electrode and driven right leg (DRL) passive electrode served as the ground. EEG data were recorded and streamed to MATLAB using the FieldTrip real-time buffer[65]. EEG data were loaded into MATLAB for template generation using the BioSig toolbox[66]. EEG epochs used in real-time and offline analyses were average referenced, baseline corrected, linearly detrended and notch filtered at 50 Hz.

To determine the optimal EEG channels for template generation and real-time analyses, averaged SSVEPs at each occipitoparietal electrode site (Iz, O1, Oz, O2, PO7, PO3, POz, PO8, PO4, P1, P3, P5, P7, P9, Pz, P2, P4, P6, P8, P10) were generated for each of the 28 letters/frequencies. These averaged SSVEPs were zero padded to 5.0 s, allowing 0.2 Hz spectral resolution, and submitted to fast Fourier transforms (FFTs). The four channels showing maximal FFT amplitudes (µV) at each fundamental frequency were retained, such that template generation and real-time classification was based only on the unique channels within this list. These multi-channel data are henceforth referred to as single-trial EEG.

Individualized Filter-Bank CCA Classification. The classification procedure determined which frequency/key was most likely selected by finding the template frequency that most strongly correlated with the single-trial EEG, as outlined in Fig. 8. The analysis evaluated the overall pattern of five weighted correlations between the three input variables: single-trial EEG (*x*; Fig. 8a), template sinusoids (*Z*; Fig. 8b), and template SSVEPs (*Y*; Fig. 8b), separately for each potential frequency (f_i) and harmonic (n_j):

$$r = \begin{bmatrix} r(1) \\ r(2) \\ r(3) \\ r(4) \\ r(5) \end{bmatrix} = \begin{bmatrix} \rho(\beta_x(xZ)x, \ \beta_z(xZ)Z) \\ \rho(\beta_x(xY)x, \ \beta_Y(xY)Y) \\ \rho(\beta_x(xZ)x, \ \beta_Z(xZ)Y) \\ \rho(\beta_x(xY)Y, \ \beta_Y(xY)Y) \\ \rho(\beta_Z(YZ)x, \ \beta_Y(YZ)Y) \end{bmatrix},$$

(3)

Where $\rho(a,b)$ represents the weighted correlation between variables a and b, and $\beta_c(c,d)$ represents the weights of c from the canonical correlation of c and d. The five elements of this vector were combined for each harmonic to form a single selection feature, preserving the sign of each weighted correlation:

$$\rho_{f_i n_j} = \sum_{i=1}^{5} sign(r(i)) \cdot r(i)^2 \tag{4}$$

The weighted sum of squares across harmonics was computed as the final feature for target frequency identification:

$$R_{f_i} = \sum_{n_j=1}^{N} \frac{1}{n_j} \cdot \rho_{f_i n_j}^2 \tag{5}$$

This classification feature (R_{fi}) was calculated for each frequency. The frequency/key for which the value was maximal was identified as the selected frequency/key (Fig. 8c).

Template Training. **Overview**. The single-trial EEG recorded during the template training phase was used to create individualised templates of the neural activity evoked by focusing attention on each key. Template signals consisted of 140 (5 harmonics × 28 frequencies) SSVEPs and 140 sinusoids matching in frequency and phase. To increase classification accuracy, the first 0.25 s were excluded, as this reflected a frequency non-specific evoked response.

Template Bank of SSVEPs. SSVEPs were constructed by averaging the single-trial EEG across the 20 cued trials for each frequency. SSVEPs were bandpass filtered to five harmonic ranges (n_{1-5}) using a 4th order Butterworth filter:

$$\begin{aligned} Hz_{highpass} &= n_j(f_1 - \Delta f) \\ Hz_{lowpass} &= n_j(f_{28} + \Delta f) \end{aligned} \, for \left\{ n_j \in \mathbb{Z}^+ \big| n_j \leq 5 \right\}, \tag{6}$$

Where Hz represents the filter frequency, n_j represents the harmonic range, f_1 represents the lowest flicker frequency, f_{28} represents the highest flicker frequency, and Δf represents the difference between adjacent frequencies.

Template Bank of Sinusoids. The procedure for generating template sinusoids matching the SSVEP frequencies/phases is outlined in Fig. 9. SSVEPs are apparent in the EEG after an onset delay (τ; Fig. 9a), which potentially varies across individuals and frequencies. The sinusoid phase was therefore calculated separately for each SSVEP frequency and harmonic. An initial bank of potential sinusoids was constructed for each frequency (f_i) at each harmonic range (n_j), including 20 distinct phases (φ_k) from 0 to 1.9π (Fig. 9b):

$$Y_{f_i n_j}(t) = \sin\left(2\pi f_i n_j t + \phi_k\right) for \{0 \, s < t \leq 1.5 \, s\} \tag{7}$$

The optimal phase for classification was chosen separately for each frequency and harmonic. To determine the optimal phase, all single-trial epochs corresponding to frequency f_i were bandpass filtered to the harmonic range n_j (Fig. 9a). The filtered epochs were correlated (using CCA) with sinusoids at all flicker frequencies and phases (0–1.9π) at harmonic range n_j (Fig. 9c). For the CCA, the sinusoid was the univariate measure and the single-trial EEG was the multivariate measure. For each filtered epoch (1–20) at frequency f_i and sinusoid phase (0–1.9π), we compared the strength of the canonical correlation (r) with each sinusoid frequency (harmonics of 10–15.4 Hz). If the maximum correlation for a given epoch was with a sinusoid at the input frequency (f_i), the corresponding sinusoid phase was scored as "correct" for that epoch. However, if the maximum correlation was with any other sinusoid frequency, the corresponding phase was scored as "incorrect" (Fig. 9d). This allowed us to determine the "accuracy" for each phase by tallying across the 20 epochs (Fig. 9e). The phase with the highest accuracy was chosen to be the sinusoid phase for frequency f_i and harmonic n_j. In summary, for each frequency and harmonic, we chose the sinusoid phase that maximised classification accuracy by correlating most strongly with SSVEPs at the input frequency during training.

Stimulus Presentation. Each key (i) was tagged with sinusoidal flicker at a unique frequency (f_i)/phase (φ_i, with luminance as a function of time (s) since flicker onset (see Fig. 1):

$$lum(t) = \frac{1}{2}(1 + \sin(2\pi f_i t + \phi_i)) \tag{8}$$

$$f_i = f_0 + \Delta f(i-1) \, for \, \{i \in \mathbb{Z}^+ | i \leq 28\} \tag{9}$$

where $f_0 = 10.0 \, Hz$, $\Delta f = 0.2 \, Hz$

$$\phi_i = \phi_0 + \Delta\phi(i-1) \, for \, \{i \in \mathbb{Z}^+ | i \leq 28\} \tag{10}$$

where $\varphi_0 = 1.5\pi$, $\Delta\varphi = 0.35\pi$

Keys subtended 3.7°[2] of visual angle, and 1.1° separated adjacent keys. Keys were outlined by a 0.1° white border. Key characters were presented in green 50 pt. Andale Mono font. During training, keys were cued by a red outline and arrow. A white textbox (43.2° × 4.0°) appeared at the top of the display. BCI typed text appeared in

Calculation of Phase Shift for EEG Epochs at Example Frequency *i* and Harmonic *j*.
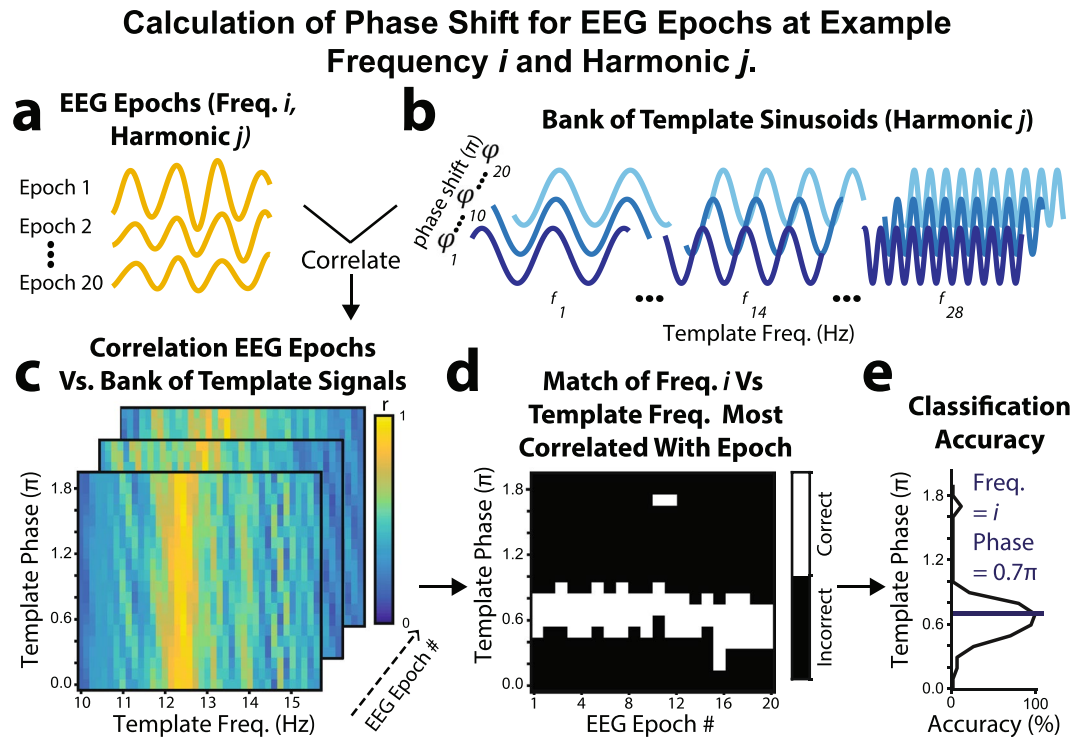
Figure 9. Procedure for generating the template bank of sinusoids with optimal classification phases, calculated separately for each frequency (*i*) and harmonic (*j*; five harmonics in total). (**a**) Single-trial EEG corresponding to cued flicker periods of template training was bandpass filtered to harmonic ranges that encompassed the lowest and highest frequencies for that harmonic (i.e., $n_j f_1 - n_j f_{28}$; e.g., $1f$: 9.8–15.6 Hz). The single trial EEG was canonically correlated with (**b**) each potential template sinusoid (phases 0.0–1.9π), resulting in (**c**) canonical *r* values representing the correlation between each single-trial epoch and template sinusoids at each phase. (**d**) For each epoch, the analysis identified the maximally correlated template sinusoid frequency. If this frequency was the cued frequency (*i*), the corresponding sinusoid phase was coded as "correct", otherwise the phase was coded as "incorrect". (**e**) The mean classification accuracy across epochs determined the optimal phase of the sinusoid at each frequency and harmonic. In this example, a sinusoid phase of 0.7π maximized classification accuracy.

this textbox (100 pt. font). During the QWERTY classification assessment and free BCI communication phases, the three last classified characters were printed at the top of each key (40 pt. font), allowing participants to keep track of their BCI typing.

Stimuli were presented at a viewing distance of 57 cm on a 24-inch ASUS VG248QE LCD monitor running at 1920 × 1080 at 144 Hz using the Cogent 2000 Toolbox (http://www.vislab.ucl.ac.uk/cogent.php) running in MATLAB R2016b (64-bit) under Windows 10 (64-bit). The computer contained an Intel Xeon E7–4809 v2 CPU and NVIDIA QUADRO M4000 GPU. The experiment was conducted in a darkened room and participants' head positions were stabilized with a chin rest. Eye tracking data were also collected for Experiment 1 but are beyond the scope of the present report, which focuses on free communication using BCI alone.

Experiment 2: Methods
Two experienced BCI participants (ages of 25 and 33 years) were instructed to have a free conversation using their brain activity (Fig. 6a). Both participants provided informed consent for both study participation and the publication of the conversation transcript in an online open-access publication. Both participants had previously completed three pilot runs in the development of the messaging interface. The two participants viewed separate displays and used BCI systems running on separate computers, with BCI typed messages sent across the local area network via TCP/IP. In addition to the 28 flickering keys of Experiment 1 (Fig. 6b), we introduced an EMG enter key [↵] based on electrical potentials evoked by jaw clenching. The enter key allowed the two participants to complete their entries and to view the sentences/phrases most recently entered by their conversational partner. A chat icon indicated whether the second participant was currently BCI typing (icon enabled) or viewing messages (icon disabled; Fig. 6c). The two participants used the enter key at will to resume typing. The EMG analysis involved FFTs performed on the most recent 1 second of data (spectral resolution: 1 Hz) recorded from the frontal scalp electrode FPz. The [↵] key was deemed selected if mean FFT amplitude (µV) in the range of 50–100 Hz exceeded the criterion of 4.0 or 4.5 µV, set separately for the two participants. A cool-down period required that successive enter keystrokes were separated by > 4 s.

The software consisted of three parallel threads devoted to reading from the amplifier, real-time analysis and stimulus presentation (Fig. 6d). Communication between threads was via parallel port (DB25) triggers and TCP/IP using the FieldTrip real-time buffer[65]. The BCI keyboard and message display were presented using the Unity game engine (Version 2018.2.8f1, Unity Technologies) running at 144 Hz and 1920 × 1080 pixels with VSync enabled on an NVIDIA GeForce GTX 1080 GPU. The FieldTrip buffer was managed in Unity using a C# API (https://github.com/georgedimitriadis/androidfieldtripbufferinunity). Template generation and real-time classification were performed using eight workers controlled via MATLAB 2017b's Parallel Computing Toolbox (64-bit) running on an Intel(R) Xeon(R) W-2145 CPU @ 3.70 GHz CPU. The calculations underlying template generation required ~2 minutes of processing. Real-time classification occurred during the flicker-free period and took on average ~259 ms (mean of $N = 1340$ classifications; $SD = 60$ ms, range: 135–408 ms).

EEG was sampled at 1200 Hz from g.USBamp amplifiers (one for each participant; g.tec Medical Engineering, GmbH, Austria) from six active gel g.SCARABEO Ag/AgCl scalp electrodes connected to a g.GAMMAbox and arranged in a g.GAMMAcap according to the international standard 10–20 system for electrode placement (Oostenveld and Praamstra, 2001). Note that different amplifiers were used in Experiments 1 and 2. This allowed us to test the two participants of Experiment 2 with identical code on identical amplifiers to each other. SSVEPs at five occipitoparietal electrodes (Iz, O1, O2, Oz, POz) were used to create training templates and for real-time classification of the attended BCI key. As noted, electrode FPz was used for the EMG enter key [↵]. The ground electrode was positioned at FCz, and the reference electrode was attached to the left earlobe via a clip. Data were band-pass (1–100 Hz) and notch (50 Hz) filtered in real-time at the hardware level. EEG signal quality was established by inspection of the real-time traces visualized by MATLAB's *dsp.TimeScope* object. The amplifiers were controlled using the g.tec NEEDaccess MATLAB API V1.16.00.

The real-time classification algorithms were identical to those of Experiment 1 (including the use of five harmonics), with the exceptions that the two participants completed 15 (rather than 20) template training blocks (420 trials; 14 minutes) and that SSVEP templates and real-time classification were based on five (rather than four) occipitoparietal electrodes. As in Experiment 1, the flicker period was 1.50 s, and the flicker-free period was 0.50 s for template training and 0.75 s for free communication. The flicker frequencies/phases were identical to those of Experiment 1. The sinusoidal luminance modulation underlying flicker was calculated based on time elapsed from the onset of the first flicker frame using the *System.Diagnostics.Stopwatch* C# class. Frame rates during flicker were confirmed to be stable at 144 Hz using Unity-recorded flip times (*Time.deltaTime*), amplifier-recorded inter-trigger spacing and photodiode measurements. Stimuli were presented on ASUS VG248QE LCD monitors at a viewing distance of 49 cm. The experiment was conducted in a darkened room, and the two participants were separated via a partition.

References

1. Sanchez-Vives, M. V. & Slater, M. From presence to consciousness through virtual reality. *Nat Rev Neurosci* **6**, 332, https://doi.org/10.1038/nrn1651 (2005).
2. Matthews, D. Virtual-reality applications give science a new dimension toolbox. *Nature* **557**, 127–128, https://doi.org/10.1038/d41586-018-04997-2 (2018).
3. Skibba, R. Virtual reality comes of age. *Nature* **553**, 402–403, https://doi.org/10.1038/d41586-018-00894-w (2018).
4. Cipresso, P., Giglioli, I. A. C., Raya, M. A. & Riva, G. The past, present, and future of virtual and augmented reality research: A network and cluster analysis of the literature. *Front Psychol*, **9**, https://doi.org/10.3389/fpsyg.2018.02086 (2018).
5. Ward, D. J. & MacKay, D. J. Artificial intelligence: fast hands-free writing by gaze direction. *Nature* **418**, 838, https://doi.org/10.1038/418838a (2002).
6. Lotte, F. *et al*. A review of classification algorithms for EEG-based brain–computer interfaces: a 10 year update. *J Neural Eng* **15**, 031005, https://doi.org/10.1088/1741-2552/aab2f2 (2018).
7. Lindgren, J. T. As above, so below? Towards understanding inverse models in BCI. *J Neural Eng* **15**, 012001, https://doi.org/10.1088/1741-2552/aa86d0 (2017).
8. DeBettencourt, M. T., Cohen, J. D., Lee, R. F., Norman, K. A. & Turk-Browne, N. B. Closed-loop training of attention with real-time brain imaging. *Nat Neurosci* **18**, 470–478, https://doi.org/10.1038/nn.3940 (2015).
9. Shibata, K., Watanabe, T., Sasaki, Y. & Kawato, M. Perceptual learning incepted by decoded fMRI neurofeedback without stimulus presentation. *Science* **334**, 1413–1415, https://doi.org/10.1126/science.1212003 (2011).
10. Lebedev, M. A. & Nicolelis, M. A. Brain–machine interfaces: past, present and future. *Trends Neurosci* **29**, 536–546, https://doi.org/10.1016/j.tins.2006.07.004 (2006).
11. Bouton, C. E. *et al*. Restoring cortical control of functional movement in a human with quadriplegia. *Nature* **533**, 247, https://doi.org/10.1038/nature17435 (2016).
12. Rezeika, A. *et al*. Brain-computer interface spellers: a review. *Brain Sci* **8**, 57, https://doi.org/10.3390/brainsci8040057 (2018).
13. Pandarinath, C. *et al*. High performance communication by people with paralysis using an intracortical brain-computer interface. *eLife* **6**, e18554, https://doi.org/10.7554/eLife.18554 (2017).
14. Beukelman, D. R., Fager, S., Ball, L. & Dietz, A. AAC for adults with acquired neurological conditions: A review. *Augment Altern Commun* **23**, 230–242, https://doi.org/10.1080/07434610701553668 (2007).
15. Birbaumer, N. *et al*. A spelling device for the paralysed. *Nature* **398**, 297, https://doi.org/10.1038/18581 (1999).
16. Cohen, M. X. Where Does EEG Come From and What Does It Mean? *Trends Neurosci*, https://doi.org/10.1016/j.tins.2017.02.004 (2017).
17. Li, L. & Negoita, S. Brain-to-speech decoding will require linguistic and pragmatic data. *J Neural Eng* **15**, 063001, https://doi.org/10.1088/1741-2552/aae466 (2018).
18. Wolpaw, J. R., Birbaumer, N., McFarland, D. J., Pfurtscheller, G. & Vaughan, T. M. Brain–computer interfaces for communication and control. *Clin Neurophysiol* **113**, 767–791, https://doi.org/10.1016/S1388-2457(02)00057-3 (2002).
19. Lotte, F., Congedo, M., Lécuyer, A., Lamarche, F. & Arnaldi, B. A review of classification algorithms for EEG-based brain–computer interfaces. *J Neural Eng* **4**, R1, https://doi.org/10.1088/1741-2552/aab2f2 (2007).
20. Zerafa, R., Camilleri, T., Falzon, O. & Camilleri, K. P. To train or not to train? A survey on training of feature extraction methods for SSVEP-based BCIs. *J Neural Eng*, https://doi.org/10.1088/1741-2552/aaca6e (2018).

21. Chen, X. *et al.* High-speed spelling with a noninvasive brain–computer interface. *Proc Natl Acad Sci USA* **112**, E6058–E6067, https://doi.org/10.1073/pnas.1508080112 (2015).
22. Waytowich, N. R. *et al.* Compact convolutional neural networks for classification of asynchronous steady-state visual evoked potentials. *arXiv preprint*, arXiv:1208.6109 (2018).
23. Akram, F., Han, H.-S. & Kim, T.-S. A P300-based brain computer interface system for words typing. *Comput Biol Med* **45**, 118–125, https://doi.org/10.1016/j.compbiomed.2013.12.001 (2014).
24. Speier, W. *et al.* A comparison of stimulus types in online classification of the P300 speller using language models. *PloS One* **12**, e0175382, https://doi.org/10.1371/journal.pone.0175382 (2017).
25. Farwell, L. A. & Donchin, E. Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. *Electroencephalogr Clin Neurophysiol* **70**, 510–523, https://doi.org/10.1016/0013-4694(88)90149-6 (1988).
26. Lin, Z., Zhang, C., Zeng, Y., Tong, L. & Yan, B. A novel P300 BCI speller based on the Triple RSVP paradigm. *Sci Rep* **8**, 3350, https://doi.org/10.1038/s41598-018-21717-y (2018).
27. Guger, C. *et al.* How many people are able to control a P300-based brain-computer interface (BCI)? *Neurosci Lett* **462**, 94–98, https://doi.org/10.1016/j.neulet.2009.06.045 (2009).
28. Krusienski, D. J. *et al.* A comparison of classification techniques for the P300 Speller. *J Neural Eng*, **3**, https://doi.org/10.1088/1741-2560/3/4/007 (2006).
29. Krusienski, D. J., Sellers, E. W., McFarland, D. J., Vaughan, T. M. & Wolpaw, J. R. Toward enhanced P300 speller performance. *J Neurosci Methods* **167**, 15–21, https://doi.org/10.1016/j.jneumeth.2007.07.017 (2008).
30. Speier, W., Arnold, C. & Pouratian, N. Integrating language models into classifiers for BCI communication: A review. *J Neural Eng* **13**, 031002, https://doi.org/10.1088/1741-2560/13/3/031002 (2016).
31. Segers, H. *et al.* In *15th Nordic-Baltic Conference on Biomedical Engineering and Medical Physics*. 164–167 (Springer).
32. Hwang, H.-J., Lim, J.-H., Lee, J.-H. & Im, C.-H. In *International Winter Workshop on Brain-Computer Interface (BCI)*. 81–83 (IEEE).
33. Lin, K., Wang, Y. & Gao, X. In *38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. 5873–5876 (IEEE).
34. Sengelmann, M., Engel, A. K. & Maye, A. Maximizing Information Transfer in SSVEP-Based Brain–Computer Interfaces. *IEEE Trans Biomed Eng* **64**, 381–394, https://doi.org/10.1109/TBME.2016.2559527 (2017).
35. Norcia, A. M., Appelbaum, L. G., Ales, J. M., Cottereau, B. R. & Rossion, B. The steady-state visual evoked potential in vision research: a review. *J Vis* **15**, 4, https://doi.org/10.1167/15.6.4 (2015).
36. Nakanishi, M. *et al.* Enhancing detection of SSVEPs for a high-speed brain speller using task-related component analysis. *IEEE Trans Biomed Eng* **65**, 104–112, https://doi.org/10.1109/TBME.2017.2694818 (2018).
37. Bin, G. *et al.* A high-speed BCI based on code modulation VEP. *J Neural Eng* **8**, 025015, https://doi.org/10.1088/1741-2560/8/2/025015 (2011).
38. Chen, X., Wang, Y., Gao, S., Jung, T.-P. & Gao, X. Filter bank canonical correlation analysis for implementing a high-speed SSVEP-based brain–computer interface. *J Neural Eng* **12**, 046008, https://doi.org/10.1088/1741-2560/12/4/046008 (2015).
39. Nakanishi, M., Wang, Y., Wang, Y.-T., Mitsukura, Y. & Jung, T.-P. A high-speed brain speller using steady-state visual evoked potentials. *Int J Neural Syst* **24**, 1450019, https://doi.org/10.1142/S0129065714500191 (2014).
40. Volosyak, I. SSVEP-based Bremen-BCI interface boosting information transfer rates. *J Neural Eng* **8**, 036020, https://doi.org/10.1088/1741-2560/8/3/036020 (2011).
41. Spüler, M., Rosenstiel, W. & Bogdan, M. Online adaptation of a c-VEP brain-computer interface (BCI) based on error-related potentials and unsupervised learning. *PloS One* **7**, e51077, https://doi.org/10.1371/journal.pone.0051077 (2012).
42. Yuan, P. *et al.* A study of the existing problems of estimating the information transfer rate in online brain–computer interfaces. *J Neural Eng* **10**, 026014, https://doi.org/10.1088/1741-2560/10/2/026014 (2013).
43. Fatourechi, M. *et al.* In *Seventh International Conference on Machine Learning and Applications*. 777–782 (IEEE).
44. Tang, Z., Wang, Y., Dong, G., Pei, W. & Chen, H. in *39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. 1934–1937 (IEEE).
45. Jones, E. *The life and work of Sigmund Freud.* (Basic Books., 1953).
46. Noyes, J. The QWERTY keyboard: A review. *Int J Man Mach Stud* **18**, 265–281, https://doi.org/10.1016/S0020-7373(83)80010-8 (1983).
47. Wolpaw, J. R., Ramoser, H., McFarland, D. J. & Pfurtscheller, G. EEG-based communication: improved accuracy by response verification. *IEEE Trans. Rehabil. Eng.* **6**, 326–333, https://doi.org/10.1109/86.712231 (1998).
48. Bochkarev, V., Shevlyakova, A. & Solovyev, V. The average word length dynamics as an indicator of cultural changes in society. *Social Evolution & History* **14**, 153–175 (2015).
49. Andersen, S. K. & Müller, M. M. Driving steady-state visual evoked potentials at arbitrary frequencies using temporal interpolation of stimulus presentation. *BMC neurosci* **16**, 95, https://doi.org/10.1186/s12868-015-0234-7 (2015).
50. Painter, D. R., Dux, P. E., Travis, S. L. & Mattingley, J. B. Neural responses to target features outside a search array are enhanced during conjunction but not unique-feature search. *J Neurosci* **34**, 3390–3401, https://doi.org/10.1523/JNEUROSCI.3630-13.2014 (2014).
51. Renton, A. I., Painter, D. R. & Mattingley, J. B. Differential deployment of visual attention during interactive approach and avoidance behavior. *Cereb Cortex*, 1–18, https://doi.org/10.1093/cercor/bhy105 (2018).
52. Clayton, M. S., Yeung, N. & Cohen Kadosh, R. The many characters of visual alpha oscillations. *Eur J Neurosci* **48**, 2498–2508, https://doi.org/10.1111/ejn.13747 (2017).
53. Scolari, M., Kohnen, A., Barton, B. & Awh, E. Spatial attention, preview, and popout: Which factors influence critical spacing in crowded displays? *J Vis* **7**, 1–23, https://doi.org/10.1167/7.2.7 (2007).
54. Varoquaux, G. *et al.* Assessing and tuning brain decoders: Cross-validation, caveats, and guidelines. *NeuroImage* **145**, 166–179, https://doi.org/10.1016/j.neuroimage.2016.10.038 (2017).
55. Thomas, E., Dyson, M. & Clerc, M. An analysis of performance evaluation for motor-imagery based BCI. *J Neural Eng* **10**, 031001, https://doi.org/10.1088/1741-2560/10/3/031001 (2013).
56. Tao, T., Yi, X., Xiaorong, G. & Shangkai, G. Chirp-modulated visual evoked potential as a generalization of steady state visual evoked potential. *J Neural Eng* **9**, 016008, https://doi.org/10.1088/1741-2560/9/1/016008 (2012).
57. D'albis, T., Blatt, R., Tedesco, R., Sbattella, L. & Matteucci, M. A predictive speller controlled by a brain-computer interface based on motor imagery. *ACM TOCHI* **19**, 20, https://doi.org/10.1145/2362364.2362368 (2012).
58. Höhne, J., Schreuder, M., Blankertz, B. & Tangermann, M. In *35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. 4185–4188 (IEEE).
59. Höhne, J., Schreuder, M., Blankertz, B. & Tangermann, M. A novel 9-class auditory ERP paradigm driving a predictive text entry system. *Front Neurosci*, **5**, https://doi.org/10.3389/fnins.2011.00099 (2011).
60. Kaufmann, T., Völker, S., Gunesch, L. & Kübler, A. Spelling is just a click away–a user-centered brain–computer interface including auto-calibration and predictive text entry. *Front Neurosci*, **6**, https://doi.org/10.3389/fnins.2012.00072 (2012).
61. Ryan, D. B. *et al.* Predictive spelling with a P300-based brain-computer interface: increasing the rate of communication. *Int J Hum Comput Interact* **27**, 69–84, https://doi.org/10.1080/10447318.2011.535754 (2010).
62. Cao, L. *et al.* A synchronous motor imagery based neural physiological paradigm for brain computer interface speller. *Front Hum Neurosci*, **11**, https://doi.org/10.3389/fnhum.2017.00274 (2017).

63. Baloh, R. W., Sills, A. W., Kumley, W. E. & Honrubia, V. Quantitative measurement of saccade amplitude, duration, and velocity. *Neurology* **25**, 1065–1065, https://doi.org/10.1212/WNL.25.11.1065 (1975).
64. Oostenveld, R. & Praamstra, P. The five percent electrode system for high-resolution EEG and ERP measurements. *Clin Neurophysiol* **112**, 713–719, https://doi.org/10.1016/S1388-2457(00)00527-7 (2001).
65. Oostenveld, R., Fries, P., Maris, E. & Schoffelen, J.-M. FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput Intell Neurosci* **2011**, 1, https://doi.org/10.1155/2011/156869 (2011).
66. Schlögl, A. & Brunner, C. BioSig: a free and open source software library for BCI research. *Comput J*, **41**, https://doi.org/10.1109/MC.2008.407 (2008).

Acknowledgements

Author contributions

A.I.R., J.B.M and D.R.P. designed the research. A.I.R. and D.R.P. performed the research and analysed the data; A.I.R., J.B.M and D.R.P. wrote the paper. All authors read and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information is available for this paper at https://doi.org/10.1038/s41598-019-55166-y.

Correspondence and requests for materials should be addressed to A.I.R.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.