

Decoding Silent Speech in Japanese from Single Trial EEGs: Preliminary Results

Yamaguchi H¹, Yamazaki T^{2*}, Yamamoto K³, Ueno S⁴, Yamaguchi A⁵, Ito T², Hirose S², Kamijo K⁶, Takayanagi H⁷, Yamanoi T⁸ and Fukuzumi S¹

¹Knowledge Discovery Research Laboratories, NEC Corporation, Shimonumabe, Nakahara-ku, Kawasaki, Kanagawa, Japan

²Department of Bioscience and Bioinformatics, Kyushu Institute of Technology, Kawazu, Iizuka, Fukuoka, Japan

³Japan Tobacco Inc., Tokyo, Japan

⁴SCSK Corporation, Osaka, Japan

⁵Hitachi Systems Ltd., Tokyo, Japan

⁶Medical Solutions Division, NEC Corporation, Tokyo, Japan

⁷Future University Hakodate, Tokyo, Japan

⁸Hokkai Gakuen University, Sapporo, Japan

Abstract

We propose a new scheme for speaker-dependent silent speech recognition systems (SSRSs) using both single-trial electroencephalograms (EEGs) scalp-recorded and speech signals measured during overtly and covertly speaking “janken” and “season” in Japanese. This scheme consists of two phases. The learning phase specifies a Kalman filter using spectrograms of the speech signals and independent components (ICs), whose equivalent current dipole source localization (ECDL) solutions were located mainly at the Broca’s area, of the EEGs during the actual speech. In case of the “season” task, the speech signals were transformed into vowel and consonant sequences, and these relationships were learned by hidden Markov model (HMM) with Gaussian mixture densities. The decoding phase predicts spectrograms for the silent “janken” and “season” using the Kalman filter with the EEGs during the silent speech. For the silent “season”, the predicted spectrograms were inputted to the HMM, and which “season” was silently spoken was determined by the maximal log-likelihood among each HMM. Our preliminary results as training steps are as follows: the silent “jankens” were correctly discriminated; the silent “season”-HMMs worked well, suggesting that this scheme might be applied to the discrimination between all the pairs of the *hiraganas*.

Keywords: Silent speech recognition system; Electroencephalogram; Kalman filter; Spectrogram; Independent component analysis; Equivalent current dipole source localization; Hidden Markov model; Gaussian mixture density

Introduction

The decipherment of human thought from brain activity, without recourse to speech or action, is one of the most attractive and challenging frontiers of modern science. In particular, silent speech recognition systems (SSRSs) enable speech communication to be needed when an audible acoustic signal is unavailable [1]. In addition to “physical” SSRSs [2-5], in the “electrical” ones, articulation may be inferred from actuator muscle signals or predicted using command signals obtained directly from the brain. Especially, the latter could be speech prosthesis for individuals with severe communication impairments. Electroencephalography (EEG) recorded during speech production attempts have increasingly yielded the decoding of phonemes and words [6,7] and artificial speech synthesizers [8,9]. However, the SSRSs using non-invasively recorded brain activity, such as scalp-recorded EEGs [10-14], functional magnetic resonance imaging (fMRI) [15] and functional near infrared spectroscopy (fNIRS) [16], had been still in the experimental stage, and limited almost to vowel recognition. Therefore, we propose a new scheme for a speaker-dependent SSRS using single-trial scalp-recorded EEGs for silent vowel recognition, and generalize to consonant one in Japanese. The scheme consists of two phases (learning and decoding ones). In order to exemplify this scheme, we carried out two experiments (Experiments I and II).

Materials and Methods

Subjects, tasks and electrical recordings

Ten healthy student volunteers (two females and eight males; mean age: 23.7 ± 1.42 years) participated in Experiment I, whose procedures

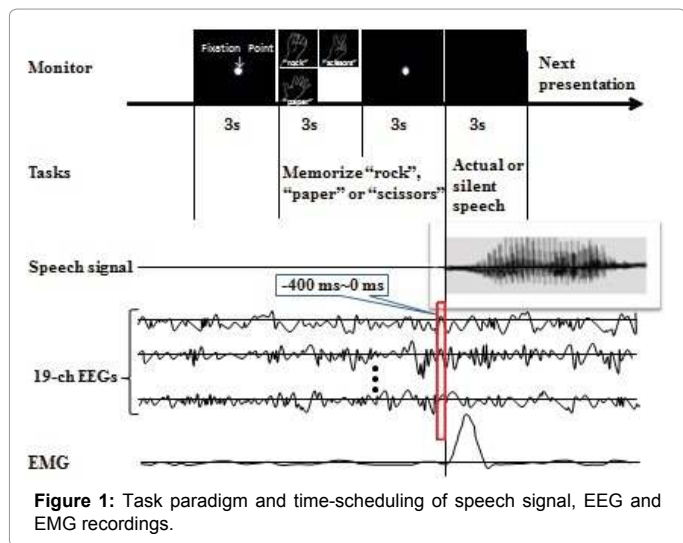
were approved by the Ethics Committee for Human Subject Research, Faculty of Computer Science and Systems Engineering, Kyushu Institute of Technology. Informed consents were obtained from all the subjects in writing for the procedures prior to the experiment. All the subjects were right-handed according to the Edinburgh inventory [17]. The subjects were requested to speak “rock”, “paper” or “scissors” (/gu:/, /pa:/ or /tʃɔki/ in English pronunciation of Japanese, respectively) into a microphone (MS-STM87SV, ELECOM CO., LTD., Japan) in the learning phase or to silently speak it in the decoding phase, according to visual cues. After the subjects gazed for 3 s a point presented at the center of a monitor 62 cm away from the subjects, a line drawing of a hand indicating “rock”, “paper” or “scissors” was presented for the next 3 s. Only the fixation point was presented for the next 3 s. Then, when the point disappeared, the subjects overtly or covertly spoke “rock”, “paper” or “scissors” corresponding to the line drawing presented just before (Figure 1). The line drawings were randomly presented ten times for each janken. Nineteen active electrodes (AP-C100-0155, DIGITEX LAB. CO., LTD., Japan) were affixed to the scalp according to the International 10-20 system. Additive six channels were included for electromyograms (EMGs) and electrooculograms (EOGs), so that

***Corresponding author:** Yamazaki T, Department of Bioscience and Bioinformatics, Kyushu Institute of Technology, Kawazu, Iizuka, Fukuoka, Japan, Tel: +81-948-29-7818; E-mail: tymzk@bio.kyutech.ac.jp

Received August 18, 2015; **Accepted** August 27, 2015; **Published** August 31, 2015

Citation: Yamaguchi H, Yamazaki T, Yamamoto K, Ueno S, Yamaguchi A, et al. (2015) Decoding Silent Speech in Japanese from Single Trial EEGs: Preliminary Results. J Comput Sci Syst Biol 8: 285-291. doi:10.4172/jcsb.1000202

Copyright: © 2015 Yamaguchi H, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.



face, mouth and eye movements were monitored. The EEGs recorded at each electrode were fed to an amplifier (Polymate AP1132, DIGITEX LAB. CO., LTD., Japan) with 10000 gain and a notch filter of 60 Hz. The amplified EEGs were sampled at a rate of 1 kHz during an epoch of 3 s preceding and 3 s following each stimulus presentation. The on-line A/D converted EEG data was immediately stored on a hard disk in a personal computer (Figure 2). Note that speech signals collected by the microphone were digitalized and, if necessary, down sampled by Audacity (a free software for recording and editing sounds: <http://audacity.sourceforge.net/>), and transformed into spectrograms by WaveSurfer (a free audio and video software: <http://www.speech.kth.se/wavesurfer/>).

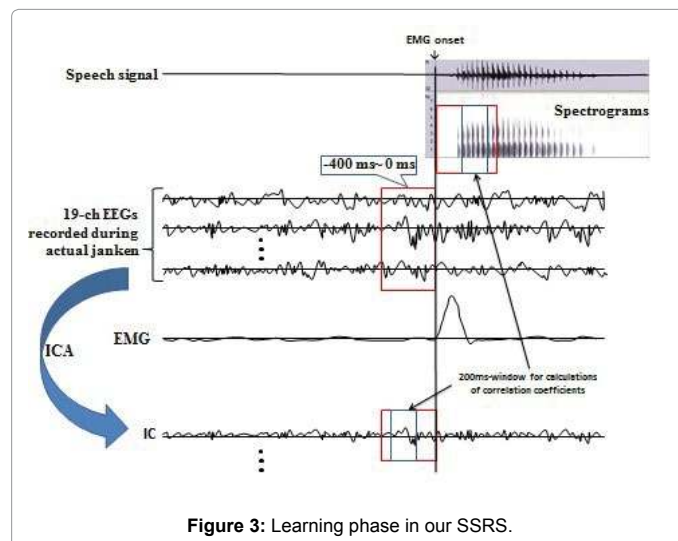
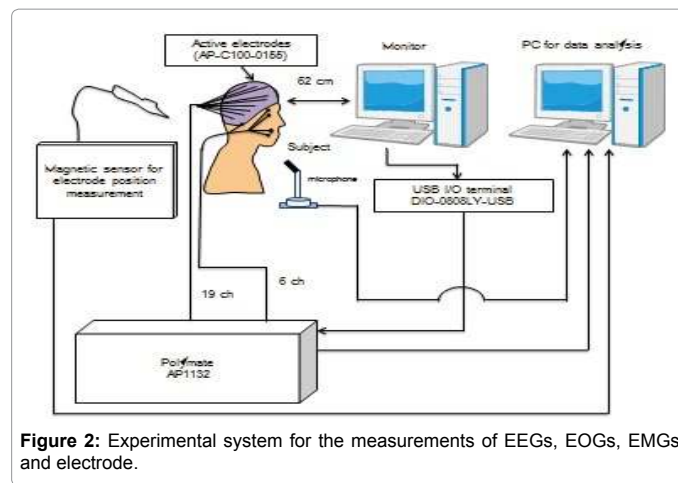
Six healthy student volunteers at the age of 23 to 27 (one female) participated in Experiment II, where a landscape photograph being associated with “spring”, “summer”, “autumn” or “winter” was presented. Task paradigm and time-scheduling of speech signals, EEGs and EMGs were the same as in Experiment I, except for 13-ch EEG recordings (F3, F5, F4, F6, F7, F8, FC3, FCz, FC4, C3, Cz, C4, POz).

Grand averages

The grand average for the actual janken was obtained by the summation time-locked to the EMG onsets of the speech signals, and one for the silent tasks time-locked to average EMG onset of the signals for each task and subject. About 900 epochs were used for these grand averages, because three sessions were carried out, each of which included 10 trials for each janken. In the latter task, the EEGs were eliminated from the averaging if the subjects overtly spoke by mistake.

ICA and ECDL

In the learning phase (Figure 3), independent component analysis (ICA) was applied to the single trial EEGs obtained. ICALAB: <http://www.bsp.brain.riken.jp/ICALAB/ICALABSignalProc/>; was used to apply the fast fixed-point ICA algorithm to the 19-ch EEGs, together with a MATLAB toolbox [18]. Then, independent components (ICs) were extracted so that their equivalent current dipole source localization (ECDL) solutions were localized to the primary motor and premotor cortices, supplementary motor area (SMA) and/or Broca’s area (BA) (Figure 4), with reference to the previous neuroimaging studies during overt articulation related to speech production [19-23].



Independent EEG sources obtained by ICA are dipolar [24]. ECDL was applied to the reconstructed EEGs, namely the projection of each of the rest ICs on the scalp surface by the deflation procedure, using “SynaCenterPro” (PC-based commercial software for multiple ECDL) (NEC corporation). This software estimates unconstrained dipoles at any timepoint [25], using the three-layered concentric sphere head model by the nonlinear optimization methods [26]. An unconstrained dipole was estimated at any timepoint with maximal peak or trough in the EEGs reconstructed by the deflation procedure for each IC.

Here, we searched for appropriate and reliable dipole solutions, by selecting localization results only with goodness of fit (GOF) of more than 90% and with the simplified confidence limits (CLs) of less than 1 mm, by restricting to the results with no drastic change in the brain sites where the unconstrained dipoles are located at least twenty successive instants including the peak or trough, and by excluding the ECDL results localized to the cerebral ventricles and the corpus callosum.

Anatomical labeling of the brain where ECDs were located, using the Japanese brain atlas for a single subject, was automatically carried out in the following: each subject’s MRI was transformed into the atlas, then the estimated ECDs were projected onto the atlas by this nonlinear transformation, and finally anatomical labels on the atlas were

determined [27]. Less electrode configurations in Experiment II were selected so that it would be easier to obtain BA-ICs, on the basis of the recent finding [28].

Kalman filter

Next, according to the hypothesis, assumed in Directions Into Velocities of Articulators (DIVA) model [8,29], that neurons in the left ventral premotor cortex present intended speech sounds in terms of formant frequency trajectories and projections from these neurons to the primary motor cortex transform the intended formant trajectories into motor commands to the speech articulators, the relationship between the extracted ICs and spectrograms of the speech signals was described by a Kalman filter. The filter was given by

$$x_t = Ax_{t-1} + w_t$$

$$A = (a_{ij}),$$

where the parameters A , C , w and v were those to be estimated [30], where x_t is the two-dimensional vector consisting of the first (F1) and second (F2) formant frequencies, y_t is the one-dimensional vector representing one IC, the matrix A describes the relationship between past and future formant frequencies, C describes the expectation of the reconstructed EEGs given a set of formant frequencies and the error terms w_t and v_t are white Gaussian random variables.

In the decoding phase (Figure 5), the inputs to the Kalman filter specified in the learning phase were the ICs whose dipole solutions were located at the premotor cortex, SMA and/or Broca's area, according to the previous neuroimaging studies related to silent speech [19,31-34], and the filter estimated spectrograms for the silent speeches using the so-called Kalman filter algorithm [35]. The 0 ms on the EEGs was defined to be average EMG onset in the learning phase for each subject (Figure 5).

HMM construction

The above SSRS for silent janken was constructed in terms of vowel recognition. Therefore, for example, spring ("haru" in English pronunciation of Japanese) and summer ("natsu" in English one) could not be discriminated by the SSRS, because the vowel transitions are the same. In order to cope with this problem, Experiment II was designated. In the learning phase for Experiment II, speech signals were transformed into vowel and consonant sequences, and these transitions were learned by hidden Markov model (HMM), in addition to spectrograms. In the decoding phase, the inputs to the HMM are spectrograms estimated from the Kalman filter specified in the learning phase. Which season was silently spoken was determined by the maximal likelihood among each HMM output in the following.

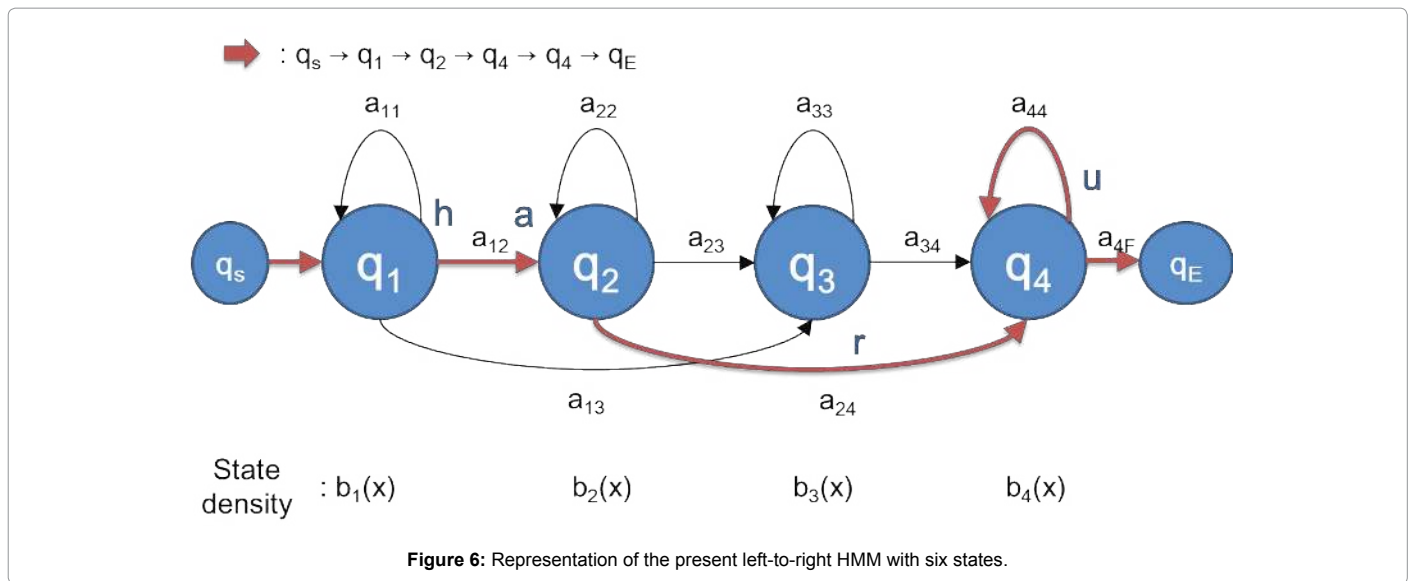
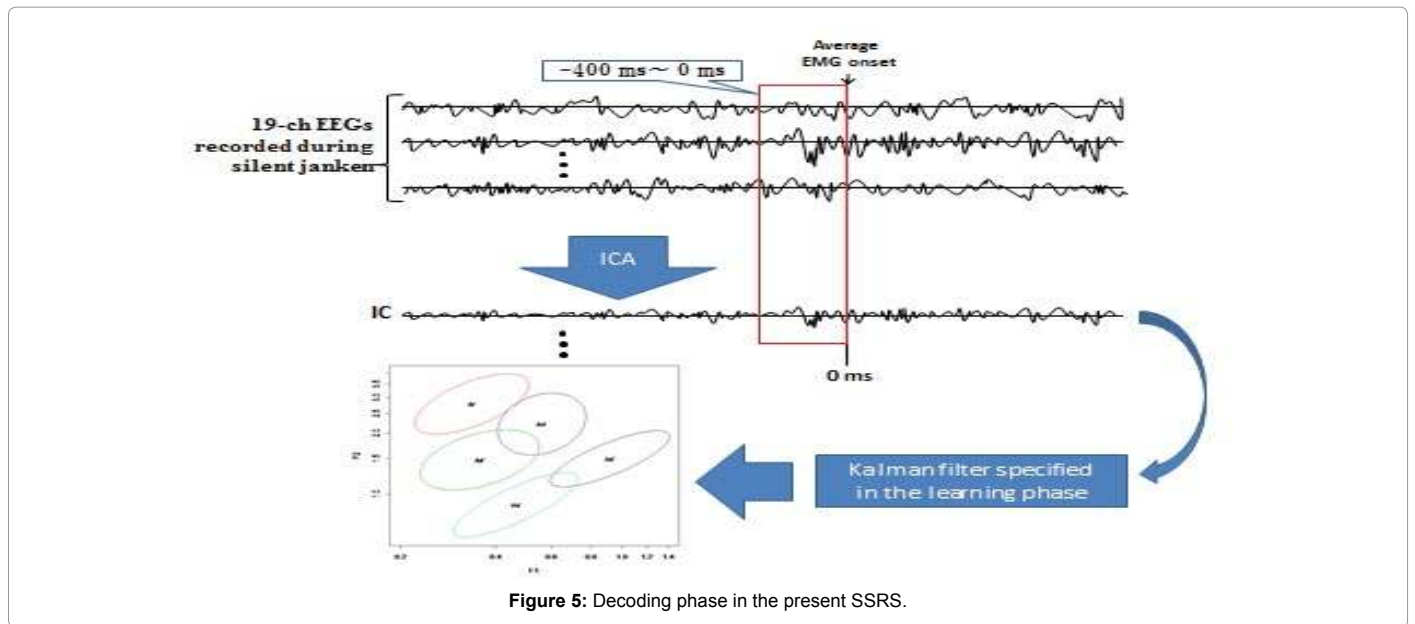
For example, the present "spring"-HMM is the left-to-right one shown in Figure 6, and is characterized by the following: (i) N is the number of states in the model; (ii) M is the number of output vowel and consonant; (iii) $A = (a_{ij})$, the transition matrix of the underlying Markov chain, where a_{ij} is the probability of making a transition from state i to state j ; (iv) $b_j(x_t)$ is the probability of outputting digitalized speech signal x_t in state j ; (v) $\pi = (\pi_i)$, $i = 1, 2, \dots, N$, the initial state probability vector. Figure 6 exemplifies $N=5$ and $M=4$. In the learning phase, each vowel and consonant occurrence is segmented into N states. This segmentation is achieved by finding the optimum state sequence, via the Viterbi algorithm with $b_j(x_t)$ modeled by Gaussian mixture densities, in addition to the initialization of π_i and a_{ij} . Parameters in the densities are estimated from spectrograms for ten actual speech trials after K-means clustering [36]. That is, for $i=1, 2, \dots, N$,

$$\log \pi_i(t=0)$$

$$f'(i,t) = \left\{ \max_j \{ f'(i,t-1) + \log a_{ij} b_j(y_t) \} \right\} (t = 1, 2, \dots, T)$$



Figure 4: A representative ECDL result for one IC extracted from one single trial EEGs, where one dipole (Blue arrow) was localized to the Broca's area.



are calculated, then a maximum log-likelihood

$$L = \max_{iq_i \in F} f'(i, T)L$$

is obtained [36], where F is a set of final states. Thus, the HMM parameters were initialized by the Viterbi algorithm and then re-estimated by the Baum-Welch algorithm. These procedures were carried out by HTK (a portable toolkit for building and manipulating HMMs in C: <http://htk.eng.cam.ac.uk/>). In the decoding phase, a silently spoken season was assumed to be maximal among each season-HMM likelihood value for the predicted spectrogram from the Kalman filter.

Results

Grand averages

Figure 7 shows the grand averages for the actual and silent janken tasks. This figure reveals similar BP-like component both for the actual

and silent speeches, while motor potential (MP)-like ones [37] for only the actual one. Therefore, for both the tasks, we should pay attention to Bereitschaftspotential (BP)-like components [38].

Correlation coefficients

We statistically examined the hypothesis in the DIVA model for our SSRS. From the correlation coefficients, with the minimal P -values, for all the tasks by all the subjects, between each formant frequency and the IC, it followed that significant correlations were found for both F1 ($r=0.62\sim 0.88$, $p \leq 6.00 \times 10^{-15}$ for “rock”; $r=0.64\sim 0.94$, $p \leq 7.64 \times 10^{-13}$ “paper”; $r=0.54\sim 0.95$, $p \leq 3.27 \times 10^{-13}$ for “scissors”) and F2 ($r=0.65\sim 0.92$, $p \leq 7.55 \times 10^{-15}$ for “rock”; $r=0.60\sim 0.98$, $p \leq 2.83 \times 10^{-11}$ for “paper”; $r=0.68\sim 0.92$, $p \leq 7.88 \times 10^{-15}$ for “scissors”) except for one subject (DK) (Table 1), thus confirming the hypothesis. The Kalman filter parameters were calculated by using one pair of the ICs and the spectrogram, whose correlation coefficient had the minimal P -value.

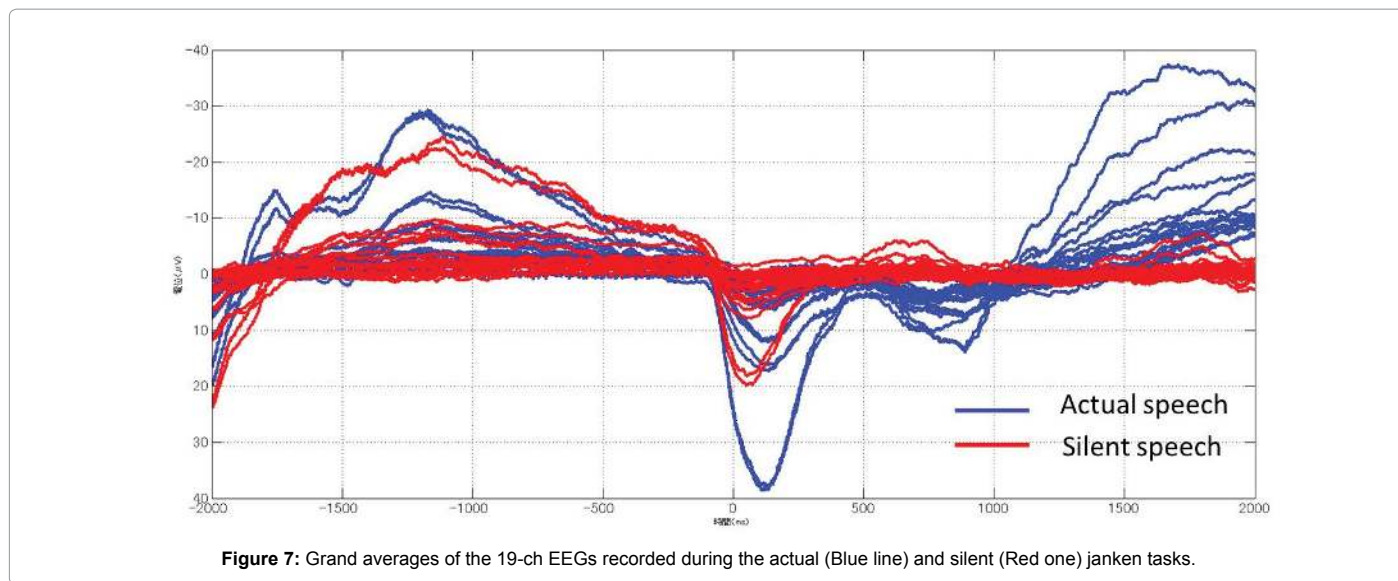


Figure 7: Grand averages of the 19-ch EEGs recorded during the actual (Blue line) and silent (Red one) janken tasks.

Rock							Paper							Scissors						
Subjects	IC1	FF1	R	IC2	FF2	r	Subjects	IC1	FF1	R	IC2	FF2	r	Subjects	IC1	FF1	R	IC2	FF2	r
DK	97	172	0.82	290	91	0.77	DK	161	175	0.85	189	143	0.85	DK	245	192	0.78	236	134	0.68
HT	271	249	0.76	264	156	0.84	HT	75	77	0.81	21	1	0.86	HT	270	257	0.81	251	251	0.88
ST	185	185	0.81	197	15	0.92	ST	297	47	0.91	205	286	0.94	ST	213	98	0.95	226	229	0.92
MS	246	62	0.75	212	207	0.78	MS	258	19	0.94	281	42	0.98	MS	118	20	0.81	204	204	0.79
KY	106	43	0.83	75	50	0.79	KY	98	61	0.88	96	231	0.94	KY	190	292	0.81	116	116	0.79
TS	183	12	0.80	233	205	0.83	TS	50	277	0.86	59	213	0.83	TS	261	34	0.79	251	251	0.82
KM	273	54	0.88	152	205	0.83	KM	208	274	0.88	143	143	0.77	KM	226	300	0.88	122	122	0.74
SS	166	301	0.62	268	301	0.69	SS	140	167	0.64	4	24	0.60	SS	174	259	0.65	35	35	0.66
HI	65	112	0.71	278	100	0.65	HI	30	7	0.79	23	265	0.77	HI	84	192	0.54	84	84	0.56
KT	149	169	0.78	1	165	0.79	KT	158	301	0.88	100	104	0.78	KT	160	11	0.86	132	132	0.70

Table 1: Correlation coefficients between each formant frequency (F1 and F2) and the IC, where the width of a time window is 300 ms, and the starting time for the calculations. IC1 or IC2 depicts the starting time (in ms) in the IC for the calculation of correlation coefficients (r), with the minimal P-values, between the IC and F1 or F2, where, in this table, 0 ms is defined to be 400 ms before the EMG onset. FF1 or FF2 represents the starting time (in ms) in the F1 or F2 for the calculation.

Estimated spectrograms for the silent janken

As the training performance, in case of one subject, diagonal parts of Figure 8 show the predicted spectrograms in the F1-F2 plane for the silent “rock”, “paper” and “scissors” with ellipsoidal distributions of five Japanese vowels [39]. In case of “rock” (/gu:/) and “paper” (/pa:/), when the formant frequency trajectories reach the /u/ and /a/ regions, respectively, the predictions were considered to be correct, while, in case of “scissors” (/tʃɔki/), the trajectory was regarded as right if it passed through the region /o (ɔ)/ then the distribution /i/. In terms of Japanese pronunciation, a main difference between “scissors” and the others is that the former has two different vowels, and the latter one. To incorporate this difference in the Kalman filter algorithm, the initial values of the covariance matrix [35] were set to be variances of F1 and F2 and their covariance. Figure 9 plots all the spectrograms for each janken in the F1-F2 plane, including the covariance matrices. Note that the covariance for “scissors” was much larger than those for “rock” and “paper”. The diagonal parts of Figure 8 shows the outputs from the Kalman filter algorithm with the initial values (V) depicted in Figure 9, and all indicates the correct predictions. The same tendencies as in Figure 8 were obtained for all the rest subjects. The rest of Figure 8 exemplifies the misapplication of our predictors. For example, the “rock” predictor correctly estimated only for the silent “rock” EEGs.

Generalization to silent consonant recognition

Table 2 shows a confusion matrix for the silent “season” tasks in terms of HMM log-likelihood values. For example, at the first row (Silent “spring” EEG), for the estimated silent “spring” spectrograms, the log-likelihood of the “spring”-HMM was higher than those of the other-HMMs. The other rows demonstrate the same tendency. Therefore, if higher log-likelihood values are accepted, it could be demonstrated that these HMMs work well. As a preliminary result, the accuracy was 86% (“spring”), 29% (“summer”), 43% (“autumn”) and 100% (“winter”) for one subject.

Discussion and Future Outlook

In order to decode silent speeches from single trial EEGs, we used Kalman filters for the vowel recognition, and HMMs for continuous speech one including consonants. The performance of the present Kalman filters might be improved in the following.

Three-dimensional kalman filter

By constructing three-dimensional Kalman filter, that is, involving F3, we obtained more discriminative results for the silent “rock” and “paper” tasks (Figure 10 (A) and (B), respectively).

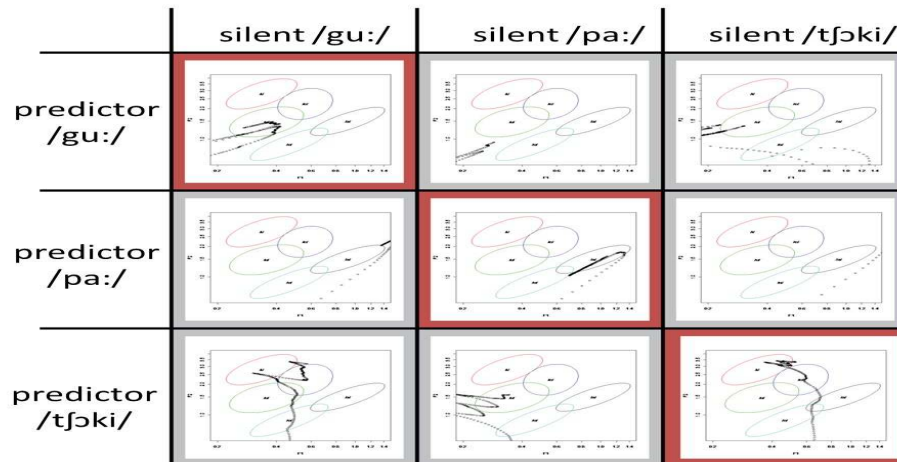


Figure 8: A confusion matrix of the three silent janken predictors for one subject.

	Silent "Spring" EEG	Silent "Summer" EEG	Silent "Autumn" EEG	Silent "Winter" EEG
"Spring"- HMM	-2844	-5749	-4663	-6441
"Summer"- HMM	-2878	-4246	-4246	-6064
"Autumn"- HMM	-3377	-5429	-4204	-6055
"Winter"- HMM	-3310	-5940	-5609	-5510

Table 2: A confusion matrix for the silent "season"-HMMs.

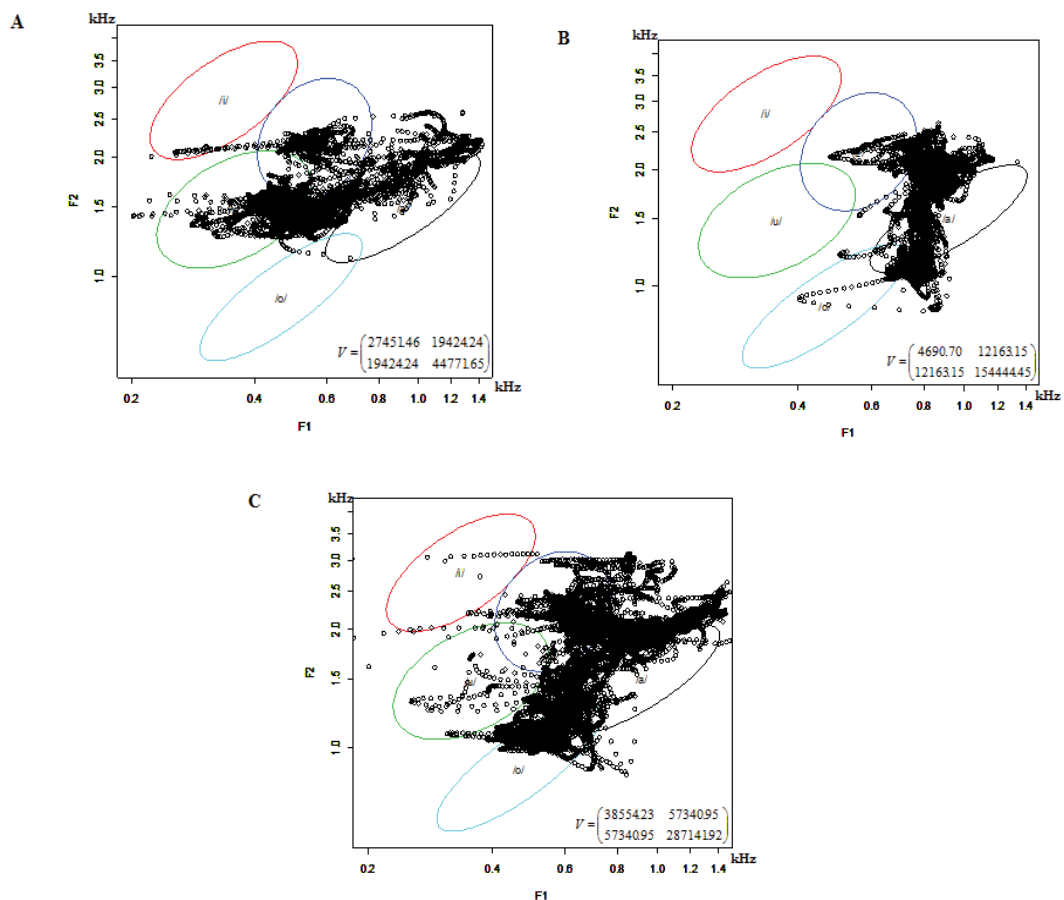
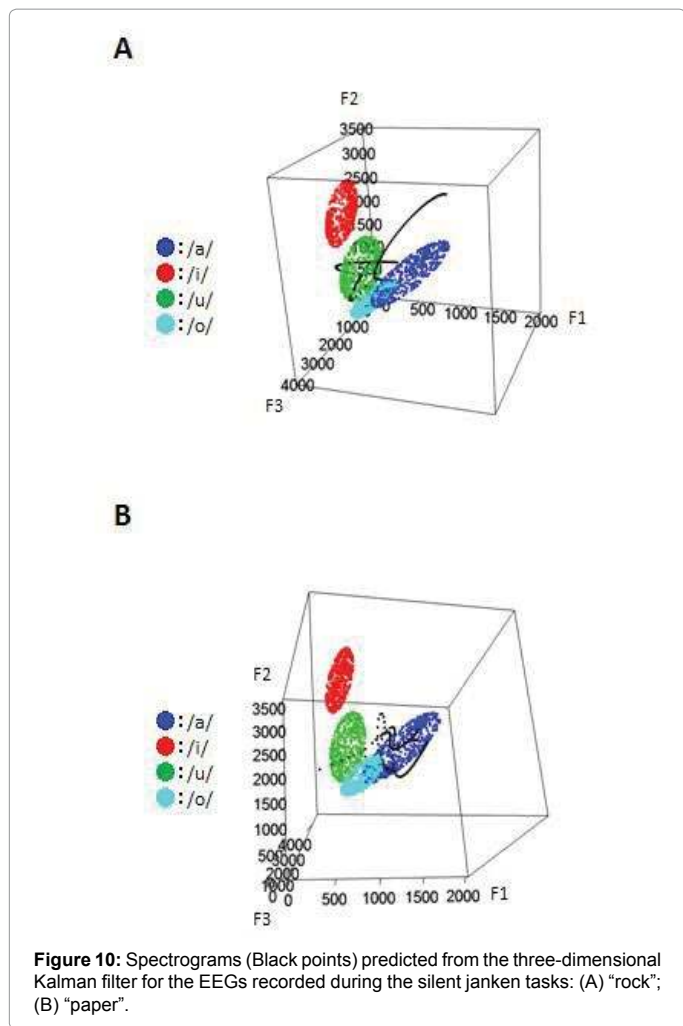


Figure 9: Spectrograms in the F1-F2 plane, obtained from the speech signals during the actual janken tasks: (A) "rock"; (B) "paper"; (C) "scissors". Each task includes 30 trials.



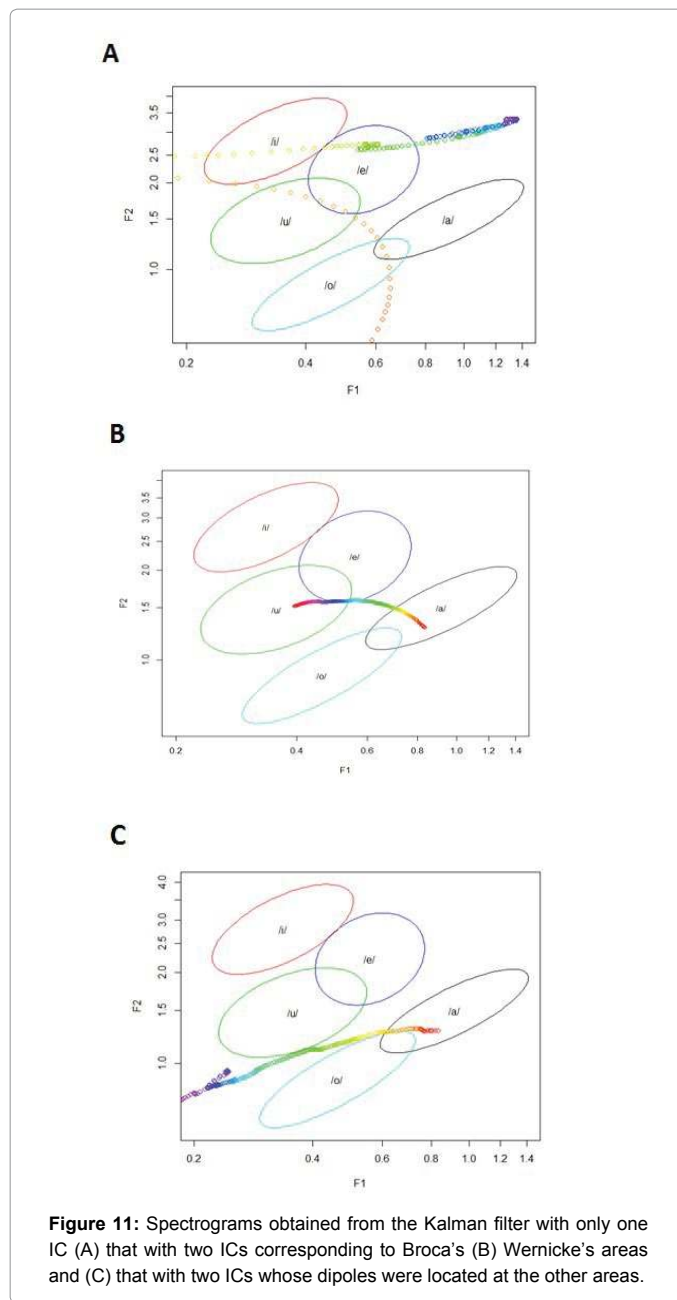
Kalman filter using two ICs

Intrinsically, vowels and consonants are known to be processed by distinct neural mechanisms [40]. For example, vowels and consonants increased activation in right middle temporal and frontal areas, respectively [41]. Tentatively, we constructed a Kalman filter with one IC whose dipole solution was located at the temporal area, in addition to the frontal-area-dipole IC, in the learning phase, and then the silent "haru" spectrogram was estimated in the learning phase. Figure 11 (A), (B) and (C) show spectrograms obtained by the Kalman filter with only one IC, that with the above two ICs and that with two ICs whose dipoles were localized to the other areas, respectively. Figure 11 (B) revealed the best performance.

Practical problem

In practice, it is unknown during which silent task EEGs were recorded. So, using all the estimated spectrograms that were obtained by all the KFs with such EEGs, all the HMMS outputted log-likelihoods. Table 3 shows a confusion matrix for one trial by one subject. This table indicates that the present method worked well.

Even if EEGs would be able to be recorded when a subject attempted to voice, speech signals might not be measured. In this case, we could obtain the speech signal by a person physically fitted to the subject



	"spring"-KF		"summer"-KF	
	"spring"-HMM "	"summer"-HMM "	"spring"-HMM "	"summer"-HMM "
silent "spring"-EEG	-3033.3	-3136.1	-3196.4	-3160.6
silent "summer"-EEG	-5819.5	-4916.0	-5018.0	-4718.4

Table 3: A confusion matrix for the "spring"- and "summer"-HMMS with respect to log-likelihoods in case of unknown silent speech EEGs.

from speech signal database (e.g., Speech Resources Consortium: <http://research.nii.ac.jp/src/en/>).

Future research

Since Japanese has syllable-timed rhythm [42], the present method could be easily generalized to all the pairs of *hiraganas*. However, because the present results are limited to the training performance,

proper testing of our method must be done in future. Moreover, there are several parameters affecting the present algorithm, such as EEG intervals which ECDL is applied to, EMG onset, speech signal features, and so on. For example, it is hoped to generalize the present framework to continuous silent speech recognition using cepstra.

Acknowledgments

This research was partly supported by a Grants-in-Aid for Scientific Research on Scientific Research (C) (23500151) and (C) (15K00276) - The Japan Society for the Promotion of Science.

References

- Denby B, Schultz T, Honda K, Hueber T, Gilbert JM, et al. (2010) Silent speech interfaces. *Speech Commun* 52: 270-286.
- Nakajima Y, Kashioka H, Shikano K, Campbell N (2003) Non-audible murmur recognition input interface using stethoscopic microphone attached to the skin. *Proceedings of IEEE ICASP* 708-711.
- Preuss RD, Fabbri DR, Cruthirds DR (2006) Noise robust vocoding at 2400 bps. 8th International Conference on Signal Processing ICSP16-20.
- Hueber T, Aversano G, Chollet G, Denby B, Dreyfus G, et al. (2007) Eigentongue feature extraction for an ultrasound-based silent speech interface. *IEEE International Conference on Acoustic, Speech, and Signal Processing ICASSP07* 1245-1248.
- Fagan MJ, Eli SR, Gilbert JM, Sarrazin E, Chapman PM (2008) Development of a (silent) speech recognition system for patients following laryngectomy. *Med Eng Phys* 30: 419-425.
- Blakely T, Miller KJ, Rao RP, Holmes MD, Ojemann JG (2008) Localization and classification of phonemes using high spatial resolution electrocorticography (ECoG) grids. *Conf Proc IEEE Eng Med Biol Soc* 2008: 4964-4967.
- Kellis S, Miller K, Thomson K, Brown R, House P, et al. (2010) Decoding spoken words using local field potentials recorded from the cortical surface. *J Neural Eng* 7: 056007.
- Guenther FH, Brumberg JS, Wright EJ, Nieto-Castanon A, Tourville JA, et al. (2009) A wireless brain-machine interface for real-time speech synthesis. *PLoS One* 4: e8218.
- Brumberg JS, Nieto-Castanon A, Kennedy PR, Guenther FH (2010) Brain-Computer Interfaces for Speech Communication. *Speech Commun* 52: 367-379.
- Callan DE, Callan AM, Honda K, Masaki S (2000) Single-sweep EEG analysis of neural processes underlying perception and production of vowels. *Brain Res Cogn Brain Res* 10: 173-176.
- Wester M (2006) Unspoken speech - speech recognition based on EEG. Master's Thesis, Universität Karlsruhe (TH), Karlsruhe, Germany.
- DaSalla CS, Kambara H, Sato M, Koike Y (2009) Single-trial classification of vowel speech imagery using common spatial patterns. *Neural Netw* 22: 1334-1339.
- Matsumoto M, Hori J (2013) Classification of silent speech using adaptive collection. *IEEE Symposium on Computational Intelligence in Rehabilitation and Assistive Technology (CIRAT)* 5-12.
- Riaz A, Akhtar S, Iftikhar S, Khan AA, Salman A (2013) Inter-comparison of classification techniques for vowel speech imagery using EEG sensors. 2nd International Conference on Systems and Informatics (ICSAI) 15-17.
- Naci L, Cusack R, Jia VZ, Owen AM (2013) The brain's silent messenger: using selective attention to decode human thought for brain-based communication. *J Neurosci* 33: 9385-9393.
- Falk TH, Paton KM, Chau T (2013) Client-centered music imagery classification based on hidden Markov models of baseline prefrontal hemodynamic responses. *Brain-Computer Interface Systems – Recent Progress and Future Prospects*. Fazel-Rezai R (ed), INTECH 137-154.
- Oldfield RC (1971) The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 9: 97-113.
- Hyvärinen A, Oja E (1997) A fast fixed-point algorithm for independent component analysis. *Neural Comput* 9: 1483-1492.
- Shuster LI, Lemieux SK (2005) An fMRI investigation of covertly and overtly produced mono- and multisyllabic words. *Brain Lang* 93: 20-31.
- Vigneau M, Beaucois V, Hervé PY, Duffau H, Crivello F, et al. (2006) Meta-analyzing left hemisphere language areas: phonology, semantics, and sentence processing. *Neuroimage* 30: 1414-1432.
- Peeva MG, Guenther FH, Tourville JA, Nieto-Castanon A, Anton JL, et al. (2010) Distinct representations of phonemes, syllables, and supra-syllabic sequences in the speech production network. *Neuroimage* 50: 626-638.
- Kielar A, Milman L, Bonakdarpour B, Thompson CK (2011) Neural correlates of covert and overt production of tense and agreement morphology: Evidence from fMRI. *J Neurolinguistics* 24: 183-201.
- Croft LJ, Rankin PM, Liégeois F, Banks T, Cross JH, et al. (2013) To speak, or not to speak? The feasibility of imaging overt speech in children with epilepsy. *Epilepsy Res* 107: 195-199.
- Delorme A, Palmer J, Onton J, Oostenveld R, Makeig S (2012) Independent EEG sources are dipolar. *PLoS One* 7: e30135.
- Mosher JC, Lewis PS, Leahy RM (1992) Multiple dipole modeling and localization from spatio-temporal MEG data. *IEEE Trans Biomed Eng* 39: 541-557.
- Kamijo K, Kiyuna T, Takaki Y, Kenmochi A, Tanigawa T, et al. (2001) Integrated approach of an artificial neural network and numerical analysis to multiple equivalent current dipole source localization. *Front Med Biol Eng* 10: 285-301.
- Tanaka K, Motoi M, Sasaguri Y, Yamazaki T, Takayanagi H, et al. (2010) P10-10 A new single-trial-EEG-based BCI – Validation of quantification methods of type II modeling. *Clin Neurophysiol* 121: S161.
- Flinker A, Korzeniewska A, Shestyuk AY, Franaszczuk PJ, Dronkers NF, et al. (2015) Redefining the role of Broca's area in speech. *Proc Natl Acad Sci U S A* 112: 2871-2875.
- Guenther FH, Ghosh SS, Tourville JA (2006) Neural modeling and imaging of the cortical interactions underlying syllable production. *Brain Lang* 96: 280-301.
- Wu W, Gao Y, Bienenstock E, Donoghue JP, Black MJ (2006) Bayesian population decoding of motor cortical activity using a Kalman filter. *Neural Comput* 18: 80-118.
- Shergill SS, Brammer MJ, Fukuda R, Bullmore E, Amaro E Jr, et al. (2002) Modulation of activity in temporal cortex during generation of inner speech. *Hum Brain Mapp* 16: 219-227.
- Keller TA, Carpenter PA, Just MA (2003) Brain imaging of tongue-twister sentence comprehension: twisting the tongue and the brain. *Brain Lang* 84: 189-203.
- Abe K, Takahashi T, Takikawa Y, Arai H, Kitazawa S (2011) Applying independent component analysis to detect silent speech in magnetic resonance imaging signals. *Eur J Neurosci* 34: 1189-1199.
- Van der Haegen L, Westerhausen R, Hugdahl K, Brysbaert M (2013) Speech dominance is a better predictor of functional brain asymmetry than handedness: a combined fMRI word generation and behavioral dichotic listening study. *Neuropsychologia* 51: 91-97.
- Kalman RE (1960) A new approach to linear filtering and prediction problems. *J Basic Eng* 82: 35-45.
- Rabiner LR, Juang BH, Levinson SE, Sondhi MM (1985) Recognition of isolated digits using hidden Markov models with continuous mixture densities. *AT&T Technical Journal* 64: 1211-1234.
- Shibasaki H, Hallett M (2006) What is the Bereitschaftspotential? *Clin Neurophysiol* 117: 2341-2356.
- Deecke L, Engel M, Lang W, Kornhuber HH (1986) Bereitschaftspotential preceding speech after holding breath. *Exp Brain Res* 65: 219-223.
- Kasuya H, Suzuki, Kido K (1968) Changes in pitch and first three formant frequencies of five Japanese vowels with age and sex of speakers. *Acoustical Society of Japan* 24: 355-364.
- Caramazza A, Chialant D, Capasso R, Miceli G (2000) Separable processing of consonants and vowels. *Nature* 403: 428-430.
- Carreiras M, Price CJ (2008) Brain activation for consonants and vowels. *Cereb Cortex* 18: 1727-1735.
- Abercrombie D (1967) *Elements of General Phonetics*. Edinburgh: Edinburgh University Press.