HHS Public Access

Author manuscript

Conf Proc IEEE Int Conf Syst Man Cybern. Author manuscript; available in PMC 2021 October 28.

Published in final edited form as:

Conf Proc IEEE Int Conf Syst Man Cybern. 2018 October ; 2018: . doi:10.1109/smc.2018.00019.

Single-Trial Classification of Disfluent Brain States in Adults Who Stutter

John C. Myers,

Department of Psychology, University of Texas San Antonio, San Antonio, United States

Farzan Irani,

Department of Communication, Disorders Texas State University, San Marcos, United States

Edward J. Golob,

Department of Psychology, University of Texas San Antonio, San Antonio, United States line

Jeffrey R. Mock,

Department of Psychology, University of Texas San Antonio, San Antonio, United States

Kay A. Robbins

Department of Computer Science, University of Texas San Antonio, San Antonio, United States line

Abstract

Normal human speech requires precise coordination between motor planning and sensory processing. Speech disfluencies are common when children learn to talk, but usually abate with time. About 5% of children experience stuttering. For most, this resolves within a year. However, for approximately 1% of the world population, stuttering continues into adulthood, which is termed 'persistent developmental stuttering'. Most stuttering events occur at the beginning of an utterance. So, in principle, brain activity before speaking should differ between fluent and stuttered speech. Here we present a method for classifying brain network states associated with fluent vs. stuttered speech on a single trial basis. Brain activity was recorded with EEG before people who stutter read aloud pseudo-word pairs. Offline independent component analysis (ICA) was used to identify the independent neural sources that underlie speech preparation. A time window selection algorithm extracted spectral power and coherence data from salient windows specific to each neural source. A stepwise linear discriminant analysis (sLDA) algorithm predicted fluent vs. stuttered speech for 81% of trials in two subjects. These results support the feasibility of developing a brain-computer interface (BCI) system to detect stuttering before it occurs, with potential for therapeutic application.

Keywords

stuttering; speech; motor control; brain-computer interface; classification

I. Introduction

Human speech requires dynamic interplay between motor control and auditory sensory systems [1]. Most adults can effortlessly map their thoughts onto lexical and acoustic representations that are then expressed as speech [2]. However, when children are learning to talk, speech production mechanisms can often fail, resulting in atypical disfluencies that are classified as stuttering [3]. Persistent developmental stuttering (PDS) is a heritable disorder characterized by disfluent speech that persists into adulthood for 1 of every 100 adults in the world population [3].

The speech of people who stutter (PWS) contains a significantly higher frequency of stuttering-like disfluencies than fluent individuals. Disfluencies such as repetitions (e.g. “Sh-sh-sh-shoe”), prolongations (e.g. “Sssssssssshoe”), and blocks (e.g. “Sh-----oe”) often occur at the beginning of an utterance, indicating problems with speech preparation [3]. This is corroborated by structural brain anomalies in speech and motor-related areas identified in PWS. Compared to fluent individuals, grey matter volume in the left inferior frontal gyrus (i.e., Broca’s area) is reduced in the brains of PWS, and the white matter in the left rolandic operculum shows reduced fractional anisotropy [4, 5]. The increased neural activity in right fronto-parietal areas during speech production (relative to controls) may reflect compensation for abnormal left hemisphere function [5, 6].

Electroencephalography (EEG) has been widely used in brain-computer-interfaces (BCIs) to classify motor imagery [7]. Measuring changes in spectral power from a pre-stimulus baseline has made it possible to classify whether a subject is imagining moving their left or right hand [7]. Just as with manual movements, speech production involves complex motor planning and stimulus processing [8]. Therefore, patterns of EEG activity might exist that similarly distinguish fluent from stuttered speech in PWS [8, 9]. Previously, we found that auditory stimuli presented during speech preparation evoke distinct neural responses in PWS compared to a fluent control group [8], with frontal beta power (~12 – 24 Hz) correlating strongly to an individual’s overall stuttering rate [9].

Our objective here is to present a method for classifying specific brain states, or patterns of neural oscillations, that support fluent speech on a single trial basis. Our strategy focuses on speech preparation because ~90% of stuttering events occur at the onset of speech [3]. We found that EEG activity during speech preparation can be used to predict whether PWS will speak fluently or stutter.

II. METHODOLOGY

A. Subject Description and Data Acquisition

Two right-handed adults (24 y/o female; 41 y/o male) who had been diagnosed with persistent developmental stuttering by a certified speech-language pathologist participated in the study. Each participant signed a consent form, and all experimental procedures were performed in accordance with a protocol approved by the University of Texas, San Antonio Institutional Review Board, consistent with the Declaration of Helsinki. EEG data was recorded using an electrode cap with 64 EEG channels (60 scalp Ag/AgCl electrodes

impedances $\leq 10 \text{ k}\Omega$) using a standard 10–20 montage. Curry 7 Neuroimaging Suite was used to digitize the data at 500 Hz with a DC–100 Hz bandpass filter (Compumedics Neuroscan, Charlotte, NC). Four electrodes were used to monitor eye movements, one above and one below the left eye and one lateral to each eye. Subjects were seated in an audiometric room in front of a computer monitor, and a pair of insert earphones was used to present auditory stimuli. Stutter trials were labeled by a speech language pathologist after each session via webcam video recordings.

B. Experimental Design

We designed the experiment to optimize measurement of spectral power changes and neuronal coherence during the period between speech preparation and production. The experimental design comprised four variations of the same ‘Get Ready-Go’ task. In each trial, subjects were presented with a ‘get ready’ stimulus (1000 ms), followed by a 1500 ms delay, and then a ‘Go’ stimulus (1000 ms) requiring a vocal response. In the primary *Word-Go (WG)* task, subjects were presented with a pseudo-word pair, followed by a 1500 ms preparatory period, and then a Go signal prompting them to speak into a microphone. The *Word-Auditory-Go (WAG)* task was similar to *WG*, except subjects were presented with a 1000 Hz pure tone (duration: 250 ms) 600 ms into speech preparation. The *Cue-Word (CW)* task required subjects to fixate on a cross at the center of the screen, followed by a 1500 ms waiting period for the pseudo-words to appear, prompting speech. Similarly, in the *Cue-Auditory-Word (CAW)* task, subjects were presented with 1000 Hz feedback between the cue and the word. A total of 800 trials were collected for each subject (2 sessions on different days, 4 tasks/session, 2 blocks/task, 50 trials/block, block-order counterbalanced). The two major factors of the design were (1) whether or not the words were presented at the beginning or end of the trial and (2) whether or not an auditory stimulus was presented between the get ready and go stimuli. The first factor examines speech preparation when subjects have either maximal or minimal lexical information before speaking (*WG* vs. *CW*). The second factor probed auditory responsiveness during speech preparation, which is known to differ in PWS and correlate to overall stuttering rate (*WAG* vs. *CAW*) [9]. Pseudo-words that mimicked words in the English language (e.g., ‘nobblesnarf globlos’) were generated using the software ‘Wuggy’ [10]. Pseudo-word pairs were used to increase stuttering frequency above the typical 10% average observed in PWS [3], allowing for a more equal proportion of fluent and stuttered trials.

C. EEG Data processing

Data were analyzed offline using the EEGLAB [11] toolbox for MATLAB (The Mathworks Inc., Natwick, MA), using an analysis pipeline summarized in Fig. 1. After average referencing, the EEG data were divided into 3-second epochs (–1000 ms to 2000 ms), time-locked to the initial word/cross presentation. All epochs were visually inspected for artifacts by looking for the abrupt onset of noise. Approximately 4% of the epochs were rejected due to eye blinks and movement artifacts. For each participant, the data were grouped by task (i.e., *WG*, *WAG*, *CW*, *CAW*) across the two sessions without normalization, because we hypothesized that the neural sources specific to each task and subject should be similar across days [12]. The data were then finite-impulse-response (FIR) filtered into two datasets that were analyzed separately for each participant: alpha-beta (8–30 Hz) and

low gamma (30–50 Hz). Filtering helps feature extraction algorithms operate within spectral ranges of interest by excluding frequency-specific artifacts such as 60 Hz line noise [12]. Although the same EEG cap was used in both sessions, the electrode positions could vary slightly across sessions.

We used the EEGLAB *runica* function with the extended option to perform an infomax independent component analysis (ICA). For time/frequency and coherence analysis, we selected the ICs that had 1) spectral peaks at alpha, beta, or gamma frequencies 2) smooth scalp projections across channels, and 3) best-fitting dipoles that mapped onto sources within the standardized Montreal Neurological Institute (MNI) head model.

D. Time Frequency and Coherence Analysis

Speech production, movement timing, and stimulus processing induce event-related spectral perturbations (ERSPs) that tend to ramp up to a maximum spectral power and ramp down across hundreds of milliseconds in a given trial [10, 11]. Let $F_{k,u}(f, t)$ denote the spectrogram of independent component u for trial k at frequency f and time t . A discrete Morlet wavelet transformation was used to compute $F_{k,u}$ with the EEGLAB *newtimef* function. Wavelets were 3 Hz for the 8–30 Hz data and 7 Hz for the 30–50 Hz data, with linear increases by 0.5 Hz for every 1 Hz increase in frequency of the input data. The average spectrogram of independent component u is given by:

$$\text{ERSP}_u(f, t) = \frac{1}{n} \sum_{k=1}^n |F_{k,u}(f, t)|^2 \quad (1)$$

where n is the number of trials in the session.

Speech preparation and production are associated with neuronal coherence [13]. Coherence is a measure of phase synchronization consistency between two IC time courses and it is theorized to capture functional connectivity between two brain regions with high temporal resolution [14]. Let $G_{uv}(f, t)$ be the average cross spectral density of independent components u and v for time t :

$$G_{uv}(f, t) = \frac{2}{n} \sum_{k=1}^n F_{ku}(f, t) F_{kv}^*(f, t) \quad (2)$$

The coherence, γ^2 , between u and v , computed with the EEGLAB function *newcrossf*, is defined as:

$$\gamma_{uv}^2(f, t) = \frac{|G_{uv}(f, t)|^2}{G_{uu}(f, t) G_{vv}(f, t)} \quad (3)$$

After computing the ERSP and coherence time series for each trial, the data were averaged across selected 200 ms time windows (as described in the next section).

E. Time Window Selection Algorithm

The features distinguishing fluent vs. stuttered trials were based on ERSP and neural coherence data computed separately on each IC time series for each of three spectral bands:

alpha (8–14 Hz), beta (15–30 Hz), and gamma (30–50 Hz). To reduce feature size and focus on salient portions of the signal, we selected 200 ms time windows centered at positions specific to each subject and variable. The total of the absolute value of the ERSP for component u in a frequency band at time t is represented by:

$$ERSP_Total_u(t) = \sum_f |ERSP_u(f, t)| \quad (4)$$

The window center, τ , for u is the time point where $ERSP_Total_u(t)$ first reaches its maximum power in the interval [0, 2000] ms. Time windows of 200 ms length were assigned around that time point for each variable (100 ms on each side of the maximum). Time window positions for each coherence variable were similarly defined for each subject.

Once the window positions were determined for each IC variable and frequency band, we computed the average value of the respective variable within its optimal window and concatenated the results into a feature vector. The average of τ over all variables for a recording was typically around 900 ms, an important time point in speech preparation for PWS [8, 9]. The selected time windows were consistent with the speech preparation phase, because speech production did occur until subjects were able to react to the go signal presented at 1500 ms in each task, nearly 2000 ms from the get ready signal. The time windows selected across frequency bands were not highly correlated (r 's < 0.20), suggesting that the properties of each IC are not redundant across frequency bands. About 20 ICs from each of the three spectral bands were selected for the analysis, resulting in large feature vectors, consisting of 400 or more variables including the ERSP of each IC and coherence between every combination of two ICs. To greatly reduce the number of variables used for classification and isolate significant contributors, we applied a stepwise linear discriminant analysis (sLDA) algorithm to the data.

F. Classification Algorithm

We used an sLDA algorithm to classify fluent vs. stuttered trials based on the ERSP/coherence data. sLDA is a robust feature reduction technique that finds a linear combination of variables that best separates two or more classes (i.e., fluent vs. stutter) [15]. Using the 'F-score to Enter' sLDA algorithm in SPSS (IBM Corp.), we maximized class separation based only on the standardized variables (z-scores) with statistically significant contributions [16]. The variables were entered in a stepwise fashion, where the highest F -score was entered first, followed by sequential entry of the next highest until all eligible variables were entered. Class separation was measured as the ratio of between-class variance to within-class variance, where greater F -scores indicated a higher probability of class separation. In the first phase of sLDA, variables with F -scores below 2.71 ($df = 1$) were automatically excluded from consideration because they represented effects with more than 10% probability of a type I error for datasets with more 120 observations across 2 classes. In the second phase, ERSP and coherence variables with F -scores above 3.84 ($df = 1$) were included because they represented effects with less than 5% probability of a type I error.

For each behavioral paradigm and subject, sLDA generated a single linear discriminant function with constant (a) whose coefficients from each predictor variable (d_n) can be fit to input data from each trial (x_n) to generate a discriminant fluency score (i.e., D):

$$D = a + d_1x_1 + d_2x_2 + d_3x_3 \dots + d_nx_n \quad (5)$$

Good classification performance is associated with greater separation between the two distributions of D for fluent vs. stuttered trials. Poor performance is associated with substantial overlap in the underlying features of each class, which would suggest that the distributions are too similar for fluent and stuttered trials to be significantly separated. As an additional defense against overfitting, we ensured that the sLDA had a minimum of 2 trials in each class per predictor variable.

III. RESULTS

A. Classification Performance

Data from both participants were subdivided into 4 parts, one subset for each behavioral task (i.e., WG, WAG, CW, CAW). The subsets were analyzed individually to assess the effects of the paradigms on the predictability of fluent vs. stuttered trials. The models for Participant 1 used a mean of 26 variables for classifying the trials in each behavioral task, and a mean of 15 variables for Participant 2.

Overall, the sLDA algorithm correctly classified 81% of trials across the two participants. Classification was accurate for 88% of fluent trials and 75% of stutter trials (see Table 1). For each paradigm, the sLDA algorithm was cross-validated by using ‘leave-one-out’ classification, which sequentially classifies each individual trial with discriminate functions derived from all other trials [15]. Results indicated that a fluent brain state can be distinguished from a disfluent brain state by using sLDA to classify the multivariate activity of neural sources (see Table 1). The models generated by sLDA for each paradigm serve as a strong starting point for improvements as more data are collected for each subject.

B. Source Localization

The time-frequency data of ICs are thought to represent spectral power changes within a compact and synchronous brain region [14]. Therefore, coherence between ICs could capture synchrony and functional connectivity between *two* brain regions [13]. The classifiers automatically selected only the linear combinations of ERSP and coherence variables that statistically separated fluent vs. stuttered trials. To better estimate whether those selected variables represent brain states that support speech, each IC projection was fitted to a single equivalent dipole using the boundary element method (BEM). BEM provides a numerical solution for realistic head geometries by forward modeling the electrical conductivity of tissues (e.g., brain, cerebrospinal fluid, skull, scalp) from segmentations of 3-D magnetic resonance images [18]. The 3-D EEG electrode coordinates were fit to the standardized MNI head model using the EEGLAB plugin *dipfit2*. Remarkably, the localized sources implicated abnormal phase synchrony between speech/motor planning and sensory regions in each behavioral task (mean residual variances (RVs) near 15%). For example, alpha coherence (8–14 Hz) between the left premotor cortex (RV: 2.73%) and the right premotor cortex (RV: 4.93%) was associated with fluent vs. stuttered trials in the WG task for Participant 1. For Participant 2, fluent vs. stuttered trials were associated with changes in beta coherence (15–30 Hz) between the right inferior frontal

gyrus (RV: 31.13%) and the left auditory cortex (RV: 12.21%). Results suggest that the time-frequency and coherence variables selected by the sLDA algorithm could operate well as direct and distinct targets for neurofeedback training.

IV. CONCLUSIONS

Here we presented a method to classify fluent vs. disfluent brain states on a single trial basis by extracting data from time windows specific to each subject and neural source. We demonstrated that neural oscillations occurring before speech attempts can be used to predict whether PWS will speak fluently or stutter. Our classification algorithm reveals some highly specific patterns of neural activity preceding fluent vs. stuttered speech. Across the two subjects, 81% of the trials were classified correctly, indicating consistency of neural signatures between sessions.

The methods summarized here have allowed us to discriminate between fluent vs. disfluent brain states in PWS. This method has an application in the much needed development of a BCI system that augments speech therapy by predicting stuttering in a neurofeedback setting. Currently, ‘fluency-shaping’ therapies have the highest efficacy in treating stuttering [19, 20]. These approaches are designed to train PWS to regulate the rhythm, onset, and tempo of their speech in order to increase fluency, but the benefits of the therapies are generally short-term [5].

With the methods presented here, classification performance in future sessions may be continually improved by adding data from each new session to a subject’s individualized training set. Real time neurofeedback will be used in conjunction with traditional speech therapy to encourage the stability of brain states favoring fluent speech to reduce stuttering severity.

References

- [1]. Hickok G (2012). Computational neuroanatomy of speech production. *Nature Reviews Neuroscience*, 13(2), 135. [PubMed: 22218206]
- [2]. Indefrey P, & Levelt WJ (2004). The spatial and temporal signatures of word production components. *Cognition*, 92(1–2), 101–144. [PubMed: 15037128]
- [3]. Bloodstein O, & Bernstein Ratner N (2008). *A handbook on stuttering* New York: Thomson Delmar Learning.
- [4]. Etchell AC, Civier O, Ballard KJ, Sowman PF (2018) A systematic literature review of neuroimaging research on developmental stuttering between 1995 and 2016. *Journal of Fluency Disorders*, 55, 6–45. [PubMed: 28778745]
- [5]. Kell CA, Neumann K, von Kriegstein K, Posenenske C, von Gudenberg AW, Euler H, & Giraud AL (2009). How the brain repairs stuttering. *Brain*, 132(10), 2747–2760. [PubMed: 19710179]
- [6]. Chang SE, Erickson KI, Ambrose NG, Hasegawa-Johnson MA, & Ludlow CL (2008). Brain anatomy differences in childhood stuttering. *Neuroimage*, 39(3), 1333–1344. [PubMed: 18023366]
- [7]. Pfurtscheller G, & Neuper C (2001). Motor imagery and direct brain-computer communication. *Proceedings of the IEEE*, 89(7), 1123–1134.
- [8]. Mock JR, Foundas AL, & Golob EJ (2015). Speech preparation in adults with persistent developmental stuttering. *Brain and language*, 149, 97–105. [PubMed: 26197258]

- [9]. Mock JR, Foundas AL, & Golob EJ (2016). Cortical activity during cued picture naming predicts individual differences in stuttering frequency. *Clinical Neurophysiology*, 127(9), 3093–3101. [PubMed: 27472545]
- [10]. Keuleers E, & Brysbaert M (2010). Wuggy: A multilingual pseudoword generator. *Behavior Research Methods* 42(3), 627–633. [PubMed: 20805584]
- [11]. Delorme A, & Makeig S (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of neuroscience methods*, 134(1), 9–21. [PubMed: 15102499]
- [12]. Delorme A, Makeig S, Fabre-Thorpe M, & Sejnowski T (2002). From single-trial EEG to brain area dynamics. *Neurocomputing*, 44, 1057–1064.
- [13]. Chen CMA, Mathalon DH, Roach BJ, Cavus I, Spencer DD, & Ford JM (2011). The corollary discharge in humans is related to synchronous neural oscillations. *Journal of cognitive neuroscience*, 23(10), 2892–2904. [PubMed: 20946054]
- [14]. Srinivasan R, Winter WR, Ding J, & Nunez PL (2007). EEG and MEG coherence: measures of functional connectivity at distinct spatial scales of neocortical dynamics. *Journal of neuroscience methods*, 166(1), 41–52. [PubMed: 17698205]
- [15]. Siddiqi MH, Ali R, Khan AM, Park YT, & Lee S (2015). Human facial expression recognition using stepwise linear discriminant analysis and hidden conditional random fields. *IEEE Transactions on Image Processing*, 24(4), 1386–1398. [PubMed: 25856814]
- [16]. IBM Corp. Released 2013. IBM SPSS Statistics for Windows, Version 22.0. Armonk, NY: IBM Corp.
- [17]. Information-based modeling of event-related brain dynamics. *Progress in brain research*, 159, 99–120. [PubMed: 17071226]
- [18]. Darvas F, Pantazis D, Kucukaltun-Yildirim E, & Leahy RM (2004). Mapping human brain function with MEG and EEG: methods and validation. *NeuroImage*, 23, S289–S299. [PubMed: 15501098]
- [19]. Webster RL (1980). The precision fluency shaping program: Speech reconstruction for stutterers. Communications Development Corporation.
- [20]. Bothe AK, Davidow JH, Bramlett RE, & Ingham RJ (2006). Stuttering treatment research 1970–2005: I. systematic review incorporating trial quality assessment of behavioral, cognitive, and related approaches. *American Journal of Speech-Language Pathology*, 15(4), 321–341. [PubMed: 17102144]

Figure 1:
EEG processing pipeline.

Figure 2:

Class separation of discriminant scores (D score) for each paradigm for two participants. Color codes (blue-fluent vs. red-stutter) were determined by the authentic group membership (i.e., ‘fluent’ vs. ‘stutter’ labels). For each data point, the y-position corresponds to the trial’s D score within its own class. The x-position corresponds to a randomly selected score from a trial within the opposite class. Model performance is reflected by point separation.

Cross validated classification performance (bold) for each subject and behavioral task. The classifiers correctly identified a majority of the original cases, thus discriminating between the brain states leading up to fluent vs. stutter speech. Note how separation improves with greater sample size in each class.

Table 1:

Participant 1	WG		WAG		CW		CAW	
	Fluent	Stutter	Fluent	Stutter	Fluent	Stutter	Fluent	Stutter
Original	79	109	131	70	81	103	72	112
sLDA %	70%	84%	98%	97%	91%	90%	93%	96%
Cross Validated %	66%	78%	92%	89%	84%	89%	88%	94%
Variance Explained	$r^2 = 0.36, p < 0.001$		$r^2 = 0.76, p < 0.001$		$r^2 = 0.67, p < 0.001$		$r^2 = 0.74, p < 0.001$	
Participant 2	Fluent	Stutter	Fluent	Stutter	Fluent	Stutter	Fluent	Stutter
	Fluent	Stutter	Fluent	Stutter	Fluent	Stutter	Fluent	Stutter
Original	151	44	174	23	147	46	146	46
sLDA %	95%	75%	98%	65%	94%	70%	95%	61%
Cross Validated %	91%	64%	97%	65%	91%	63%	93%	59%
Variance Explained	$r^2 = 0.47, p < 0.001$		$r^2 = 0.43, p < 0.001$		$r^2 = 0.48, p < 0.001$		$r^2 = 0.47, p < 0.001$	