



HAL
open science

Different encoding of legal and illegal speech sequences: beyond phonetic planning?

Anne-Lise Jouen, Cécile Fougeron, Marina Laganaro

► **To cite this version:**

Anne-Lise Jouen, Cécile Fougeron, Marina Laganaro. Different encoding of legal and illegal speech sequences: beyond phonetic planning?. *Language, Cognition and Neuroscience*, 2024, 39 (1), pp.1-15. 10.1080/23273798.2023.2241580 . hal-04177745

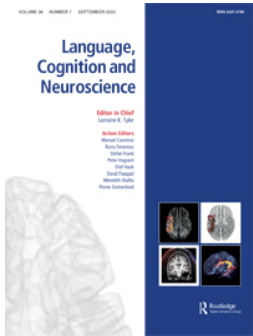
HAL Id: hal-04177745

<https://hal.science/hal-04177745>

Submitted on 11 Jan 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Different encoding of legal and illegal speech sequences: beyond phonetic planning?

Anne-Lise Jouen, Cécile Fougeron & Marina Laganaro

To cite this article: Anne-Lise Jouen, Cécile Fougeron & Marina Laganaro (2023): Different encoding of legal and illegal speech sequences: beyond phonetic planning?, *Language, Cognition and Neuroscience*, DOI: [10.1080/23273798.2023.2241580](https://doi.org/10.1080/23273798.2023.2241580)

To link to this article: <https://doi.org/10.1080/23273798.2023.2241580>



© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



[View supplementary material](#)



Published online: 31 Jul 2023.



[Submit your article to this journal](#)



Article views: 40



[View related articles](#)



[View Crossmark data](#)

Different encoding of legal and illegal speech sequences: beyond phonetic planning?

Anne-Lise Jouen^a, Cécile Fougeron^b and Marina Laganaro^a

^aFaculty of Psychology and Educational Sciences, University of Geneva, Geneva, Switzerland; ^bLaboratoire de Phonétique et Phonologie, UMR 7018, CNRS/Univ. Sorbonne Nouvelle, Paris, France

ABSTRACT

Transforming linguistic codes into articulated speech is thought to rely on different phonetic (motor speech) encoding/planning processes for practiced sequences and for unpracticed/uncommon speech sequences. However, transforming phonological codes into articulation likely involves processes beyond phonetic planning, going on even during articulation. Here we sought behavioural, acoustic and brain dynamics differences in the preparation of matched common/legal and uncommon/illegal speech sequences in 20 participants. Illegal syllables were initialised faster – contrary to what is generally expected –, had longer acoustic duration and differed from legal syllables in ERP waveform amplitudes and microstates in a time-window preceding and following the vocal onset. The pattern of results suggests that speech plans are of different size for legal and illegal syllables, and impact on the parametrisation of the corresponding motor programmes, allowing fast execution of the segmentalised illegal sequences for which incremental speech programming continues during articulation.

ARTICLE HISTORY

Received 23 December 2022
Accepted 19 July 2023

KEYWORDS



Speech production; syllable legality; speech routine; EEG; ERPs


Introduction

Speech production involves the transformation of a linguistic code into articulatory movements generating an acoustic signal. How sequences of abstract linguistic (phonological) forms are implemented into motor activities has been modelled in the framework of different theoretical approaches from different disciplines (see Parrell et al., 2019). In speech production models (Guenther, 2016; Levelt et al., 1999), speech plans are not generated on the fly each time the speaker articulates speech sequences but are rather implemented as motor routines that are retrieved from memory. The speech motor plans are syllable-sized in most models, although some models also include other possible sizes of speech plans. Crucially, as for other motor sequences, only speech sequences that have been highly practiced by the speaker of a given language are expected to generate stored routines. On the contrary, speech sequences that are not practiced enough – i.e. uncommon speech sequences – are thought to be assembled. As it will be further clarified below, uncommon speech sequences can refer either to rarely used but phonotactically legal

sequences or to speech sequences that are phonotactically illegal in a given language.

The hypothesis of different encoding processes for common and uncommon speech has been investigated with behavioural and with brain imaging approaches, leading to somehow converging results on different computational costs for practiced and unpracticed speech sequences (usually syllables). These results have mostly been interpreted in the framework of psycholinguistic models, positing a single encoding process between linguistic codes and motor execution (called “phonetic encoding” in Levelt, 1995; Levelt et al., 1999). In such models, the preparation of speech plans based on stored syllables has lower computational cost (and is presumably faster) than the building of speech plans made from assembled speech elements. Other speech production models have further expanded and detailed the transformation of an abstract linguistic code into articulation by including multiple encoding processes (Guenther, 2016; Kröger et al., 2009; Van Der Merwe, 2021). Consequently, the encoding of highly practiced versus uncommon/unpracticed speech sequences may differ at one or more of these processing levels. In the following, we aim

CONTACT Marina Laganaro  Marina.Laganaro@unige.ch  Faculty of Psychology and Educational Sciences, University of Geneva, 40 Bd du Pont d'Arve, 1205 Geneva, Switzerland

 Supplemental data for this article can be accessed online at <https://doi.org/10.1080/23273798.2023.2241580>.

© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

at (i) reviewing and reinterpreting the existing evidence for a differential encoding of common and uncommon speech sequences in the framework of multiple encoding processes and (ii) seeking behavioural, acoustic and brain dynamics evidence that differences in the encoding of common and uncommon speech sequences, which are usually ascribed only to phonetic planning, may also impact subsequent encoding processes.

Speech routines

Most speech production models assume that the transformation of linguistic codes into motor execution largely relies on overlearned coordinated speech movements that are retrieved from memory, at least for frequently used speech units (i.e. speech segments or speech sequences). These wholly stored speech routines are acknowledged in different frameworks, despite different labellings (“phonetic plans” and “mental syllabary” in Levelt et al., 1999; “speech sound maps” in the DIVA model, [Guenther et al., 2006]; “phonetic maps” in Kröger et al., 2009).

The retrieval of speech plans stored in memory is based on the idea – shared with general motor control models – that extensive (speech) practice leads to the emergence of motor routines. Motor routines involve higher-order representation of coordinated motor patterns that are accessed as a whole and do not need to be assembled each time they are needed, thus reducing the computational load of motor preparation for speaking. The stored motor details and the size of stored speech units vary across models. In psycholinguistic models, motor routines are syllable-sized (Levelt, 1995; Levelt et al., 1999). The proposal of syllables as functional speech units is based on acoustic-articulatory properties (the syllable is the main size-pattern of coarticulation and of articulatory organisation; see Krakow, 1999) and on distributional observations (a limited set of syllables is used to compose most words in a language, see Schiller et al., 1996). In neurocognitive models (Guenther et al., 2006; Kröger et al., 2009), motor routines arise as activation patterns of a trained neural network. In such models (see for instance the DIVA model), smaller (phonemes) and larger (words, phrases) units can also be stored in memory with syllables, if they are sufficiently trained. Other accounts also consider the representation of sub-syllabic components larger than phonemes (onsets, rimes, Ziegler, 2009).

As speech routines arise as a consequence of practice, it is expected that only highly practiced speech sequences are stored and retrieved from memory (Levelt & Wheeldon, 1994) while, for uncommon speech sequences, the speech plans need to be assembled

from smaller bits and this assemblage is supposed to be more costly (i.e. slower). This prediction has first been addressed with psycholinguistic behavioural studies through the comparison of initialisation speed of syllables according to their frequency of use. Faster initialisation (measured at the acoustic onset of speech after presentation of the stimulus) for frequent syllables over less frequent syllables has been reported in several languages, using immediate production tasks (Dutch: Cholin et al., 2006; Levelt & Wheeldon, 1994; French: Laganaro & Alario, 2006; English: Cholin et al., 2011). The results of better performance for frequently used syllables are also confirmed by reduced speech error rates in brain-damaged speakers suffering from apraxia of speech (Aichert & Ziegler, 2004; Laganaro, 2005; Laganaro et al., 2012; Staiger & Ziegler, 2008). However, neuroimaging studies did not find different brain areas when contrasting the production of syllables according to their frequency (Brendel et al., 2011; Carreiras et al., 2006; Papoutsi et al., 2009). These results on syllable frequency (contrasting with the results on novel or phonotactically illegal syllables that will be discussed further below) suggest that the speech plans of less frequent syllables trigger a slower access or a slower building/assemblage, but that their planning is nevertheless supported by the same brain areas. In line with this interpretation, electroencephalographic (EEG) event-related potential (ERP) studies reported different dynamics of the same microstates for the production of high versus low frequency syllables (Bürki et al., 2015, 2020).

The results summarised so far therefore raise the following questions. First, which speech preparation processes are responsible for the observed behavioural differences between frequent and infrequent syllables, given that they seem to be underpinned by the same brain networks? Second, what is meant by frequent (highly practiced) and by infrequent speech sequences? These two issues will be further exposed below.

Which speech preparation processes yield the reported behavioural effects related to the frequency of speech sequences?

Disentangling the processes sustaining these differential behavioural results for frequent/common versus infrequent/uncommon syllables is strongly related to the experimental paradigms and manipulations. As speaking involves planning both at an abstract phonological level and at phonetic/motor level(s), experimental investigation targeting phonetic/motor preparation needs to make sure that the observed results are not due to linguistic processes. For that reason, sequences of meaningless speech (usually non-words or pseudo-words) are

usually used in such studies. However, even when eliciting the production of meaningless speech, one needs to experimentally separate phonological and orthographic processes (when the pseudo-words are triggered with a written cue) from phonetic/motor speech preparation. Delayed production tasks, in which speakers are presented a speech target (orthographically or audiovisually) and then have to hold their production till they are prompted to do so, have been used to target phonetic/motor speech preparation in behavioural (Laganaro & Alario, 2006) and in neuroimaging studies (Bohland & Guenther, 2006; Chang et al., 2009; Lancheros et al., 2020; Tilsen et al., 2016). The idea behind delayed production is that the speakers can encode the phonological/orthographic content of the target speech sequences during the delay, and launch phonetic/motor encoding after the prompt triggering speech production. However, it has been claimed that phonetic/motor encoding can also be prepared during the delay (see rationale in Laganaro & Alario, 2006). For this reason, in some experimental conditions (delayed production with articulatory suppression) the participants are asked to articulate a different sequence (usually a repeated syllable) during the delay between the presentation of the target speech sequence and the prompt. This repetitive uttering task is meant to disable phonetic/motor encoding, i.e. the preparation of the targeted speech plans. In Laganaro and Alario (2006), syllable frequency effects have been reported in such conditions of delayed production with articulatory suppression, but not in standard delayed production. The frequency effect thus occurs at the level of phonetic/motor encoding, where the preparation of the targeted speech plan is impaired by the planning and execution of a distracting speech plan (the one of the articulatory suppression). It should also be reminded here that the locus of the effect is also corroborated by the syllable frequency effect observed in participants with apraxia of speech, as reported above, which underlying impairment has been ascribed to the preparation of phonetic plans (Code, 1998; Darley et al., 1975; Ziegler, 2009). Hence, the syllable frequency effect observed in speech production studies has been attributed to the preparation of speech plans in terms of “phonetic encoding” as conceptualised in psycholinguistic models positing a single process transforming the linguistic code into a motor code. Other models propose more complex motor preparation and posit distinct processes and brain circuits for the “higher-order” phonetic planning of speech plans and for the “lower-level” motor programming of these plans leading to their execution (Guenther et al., 2006; Guenther, 2016; Kröger et al., 2009), a distinction that is also routed in the clinical literature of motor speech

disorders (Duffy, 2019; Van Der Merwe, 2021). In such models, speech planning and programming encompass different processes (whereas those terms are used interchangeably in other models). Although the exact processes involved in phonetic planning and motor programming are far from being grasped, phonetic planning would entail the encoding of context- and language – specific phonetic details on the realisation of the speech goals in terms of speech units (like coordinated gestures that are somehow abstract and not muscle-specific), while speech motor programming would parametrise the motor commands needed for the execution of articulatory movements. In such frameworks, phonetic speech plans can be made of wholly stored routines (as hypothesised for overlearned speech sequences) or assembled speech units, which are then coded into detailed muscle-specific programmes conveying information on a range of movement parameters (tone, velocity, force, etc.). In the framework of a model with a planning and a programming level, it is unclear whether the syllable frequency effect is solely due to differences at the phonetic planning level (retrieval vs. assemblage into a syllable-size plan) or if differences propagate also at the programming level. Indeed, previous unexpected results of the syllable frequency effect, that remained unexplained, are compatible with an impact beyond phonetic planning (see for instance the near-to-significant differences between frequent and infrequent syllables in standard delayed production, both on behavioural responses and on brain activations in Bürki et al., 2015).

What are uncommon speech sequences: frequency vs. legality effects?

As reminded above, the hypothesis of stored/retrieved versus assembled speech plans is based on the idea that only motor behaviours that are highly practiced generate stored motor routines. Thus, most experimental manipulations compared syllables that are frequent in the target language to less frequent syllables (e.g. above 1000 occurrences per million syllables versus below 150 occurrences per million syllables in French in Laganaro & Alario, 2006; from 5.98 to 774.24 occurrences per million words versus from zero to 8.10 occurrences per million words in Cholin et al., 2011). However, given that humans speak several hours per day, it seems likely that even syllables that are less frequent at the level of the lexicon, are practiced enough to generate motor routines. By contrast, phonotactically illegal syllables (i.e. sequences of sounds that are unattested in a language as a possible syllable) are more unlikely to generate syllable-sized stored routines. Indeed, studies

comparing illegal syllables to frequent legal syllables (Bürki et al., 2015; Moser et al., 2009; Segawa et al., 2015) have reported different brain activations for the production of these two types of speech sequences (a difference that was not found when comparing frequent to infrequent syllables, see above). Interestingly, Segawa et al. (2015) and Bürki et al. (2015) also reported different patterns for legal and illegal syllables in a standard delayed production task, in which phonetic planning can be completed during the delay (although only marginally significant in Bürki et al., 2015).

It seems therefore that the contrast between legal and illegal syllables is better suited than the frequency contrast for the investigation of different encoding processes. In addition to being infrequent, “illegal” syllables are constituted of phonotactically illegal sequences in a given language, they can thus not be stored as a whole, and it is unknown whether illegal sequences are planned into a single speech plan. There is also phonetic evidence that the typical timing pattern for legal syllables is not found in the production of illegal syllables. For instance, English speakers mistime the consonants with insufficient temporal overlap in phonotactically illegal CC sequence, relative to legal English CC clusters (Davidson, 2006; Davidson & Stone, 2003).

Based on these observations, the following hypothesis can be made: any difference in preparing the production of highly practiced versus uncommon speech sequences may not be due solely to different encoding of speech plans (phonetic encoding) but may occur also during motor programming and may extend over articulation. Indeed, if speech plans for legal and illegal syllables are of different size (syllable-size vs. segment-size for instance), the parametrisation of the corresponding motor programmes should also differ as it should be implemented on units of different sizes. Here, we therefore used a standard delayed production task of legal and illegal speech sequences to seek behavioural (production latencies), acoustic (durations of the produced sequences) and EEG-ERP patterns that are compatible with an interpretation of a different encoding of these speech sequences beyond phonetic planning. We contrasted syllables composed of phonotactically legal consonant clusters in French to speech sequences composed of sequences of consonants which cannot form a legal onset in word initial position in French.

Material and methods

Participants

Previous studies using delayed production tasks with frequent/legal and infrequent/illegal syllables and

combining behavioural and brain imaging (either EEG/ERP or fMRI) approaches (see Introduction section) included from 13 to 20 participants; we therefore aimed at enrolling at least 20 participants. 25 neurotypical adults participated in the experiment (10 men; mean age: 24.8, SD = 4.7 years). They were all French native speakers with normal or corrected-to-normal vision. All subjects gave their informed consent to participate in the study, approved by the local ethics committee (NO PSE.20171103.06), and were paid for their participation. All the participants completed the task with an accuracy >75%, but 5 were excluded due to over-noisy EEG recordings and insufficient epochs in each condition, thus leaving 20 participants for the analyses.

Material

The stimuli were 112 monosyllabic pseudowords, all with an initial consonant sequence split into 2×2 matched conditions: legality (phonotactically legal vs illegal syllable in French) and structure of the sequence in which the position of the C–C sequence is either at the beginning of the word (C_1C_2) or as C_2C_3 with C_1 always being an /s/ (sC_2C_3). An example of stimuli is provided in Table 1 and the full list in the Appendix A. The control of the position of the CC consonant sequences is motivated by the fact that mixing up positions – as done in previous studies (Bürki et al., 2015, 2020; Segawa et al., 2015) – may blur the results. Finally, we also constrained the phonetic categories of the onset phonemes in the C_1C_2V condition, because different phonemic categories are known to have different articulatory to acoustic delays, which introduce undesirable variability both in the alignment of vocal onset and of response-locked ERPs, as demonstrated by (Jouen et al., 2021).

In the C_1C_2V sequences, C_1 is always a voiceless stop consonants (/p/, /t/, /k/), and it is followed either by a liquid (/l/ or /R/) in order to form a legal syllable onset in French, or by /n/, /m/, /p/, /t/, /v/, or /z/ in order to form an illegal syllable onset according to French phonotactics. The sC_2C_3V words are built with the same principles: the same CCV sequences are just preceded by the voiceless fricative /s/, thus generating either a legal

Table 1. Conditions (legality \times structure) with examples of matched stimuli (see Appendix A for the full list).

Legality	Structure	
	C_1C_2V	sC_2C_3V
legal	e.g. /kRa/, /plo/, /tRy/	e.g. /skRa/, /splo/, /stRy/
illegal	e.g. /kna/, /pmo/, /tvy/	/skna/, /spmo/, /stvy/

sC_2C_3 syllable onset or an illegal one. The legal syllables were all existing syllables in French with mean frequency of 421 per million syllables (810 and 32 for CCV and $sCCV$ respectively in the Lexique database, New et al., 2004); the illegal syllable did not exist in the French database. Stimuli were presented in a written form which matched the targeted phonetic form.

28 additional legal and illegal CCV and CCCV pseudo-words with the same initial phonemes than the ones of the target stimuli were added as filler items for the no-production condition (see Procedure section).

Procedure

The production was elicited with a delayed production task. Participants sat in front of a computer screen (approximately 70 cm) in a sound-proof dimly lit room. The experimental software E-prime (version 2.0; Schneider et al., 2002) was used for stimuli presentation and recording of audio files. First the participants were familiarised with all the pseudowords randomly presented auditory and visually, that they had to repeat overtly after each presentation in presence of an experimenter. Then, participants underwent a training phase on the delayed production task (five warm-up trials, repeated if necessary). Finally, the experimental phase started.

A trial started with a fixation cross presented for 500 ms (in white on a black screen), then a written syllable appeared on the screen and remained for 1000 ms, followed by “...” in white, which randomly lasted either 1000, 1300 or 1600 ms. A variable delay was used so that participants could not anticipate the response cue (see Laganaro & Alario, 2006 for the rationale behind the duration of the chosen delays). Filler items, for which no production was expected, were presented at the shortest delay (1000 ms). Participants were instructed to wait silently until a question mark appeared but were authorised to blink during this delay. After a brief blank screen (100 ms), the response cue (a yellow question mark) remained on the screen for 1500 ms indicating that participants had to utter the target stimulus as fast and accurately as possible. When “...” appeared in yellow instead of the question mark, participants only had to wait until the next trial. This condition was associated only with the filler items and was introduced to prevent anticipation of response on the target items.

Each item was presented twice throughout the task, once in each of the 2 delays (1300 or 1600 ms) and each filler item was presented once (252 trials in total). Items were pseudo-randomised such that the same stimulus was not presented consecutively and the

same delay was not presented for more than three consecutive trials. The task was divided in three blocks to allow participants two brief breaks in between.

EEG acquisition

The EEG signal was recorded continuously using the Active-Two Biosemi EEG system (Biosemi V.O.F., Amsterdam, Netherlands) with 128 channels covering the entire scalp. Signals were sampled at 512 Hz (filters: DC to 104 Hz, 3 dB/octave slope). The custom online reference of the system is the common mode sense – driven right leg (CMS-DRL).

Preprocessing and analyses

Production latencies (or initialisation latencies, corresponding to the lag from the question mark to the vocal onset) were extracted manually based on the display of the visualisation of the waveform and the spectrogram of each production (using CheckVocal 2.2.6; Protopapas, 2007). The vocal onset for the initial voiceless stops in the C_1C_2V stimuli corresponded to the release of closure, since no acoustic signal is present during the silent closure. For the sC_2C_3V stimuli, the vocal onset corresponded to the onset of the frication, as shown by an aperiodic signal and turbulent noise on the spectrogram. Acoustic duration of the uttered stimuli was computed from the vocal onset to the acoustic offset of the vowel. The acoustic duration of /s/ in the sC_2C_3V stimuli was computed from the apparition of high frequency turbulent noise in the acoustic signal to the burst of the following voiceless stop consonant, in order to determine the time-window of the ERP analyses during articulation (see below).

No-responses, errors (i.e. production of a different stimulus than the target), hesitations and/or auto-corrections, as well as production latencies below 200 ms and beyond 1100 ms were considered errors (either anticipations or too long latencies) and were discarded from all analyses. In addition, initialisation times and acoustic duration data were cleaned by removing observations beyond 2 SD. The remaining data (4364 observations for production latencies and 4339 for duration) were fitted with mixed models (Baayen et al., 2008) with the R-software (R-project, R-development core team 2005, version 4.2.1) with the packages *lme4* (Bates et al., 2015) and *lmerTest* (Kuznetsova et al., 2017), using backward elimination of fixed and random effects. The legality (phonotactically legal vs illegal) and the structure of the stimulus (C_1C_2V or sC_2C_3V) were entered as fixed factors along with the absolute order of the stimuli; participants and stimuli were entered as random factors. The final models are presented in the result section.

EEG pre-analyses

All the pre-processing (cleaning, epoch extraction and averaging) were computed for each participant using the Cartool software (Brunet et al., 2011). Offline, EEG data was high-pass filtered at 0.2 Hz and low-pass filtered at 30 Hz with a 2nd order acausal Butterworth filter with -12 dB/octave roll-off and notch-filtered (50 Hz). The analyses of interest were locked to the onset of the vocal (acoustic) onset (response-locked), but for the completeness of the analyses ERP epochs were also extracted locked to the stimulus onset (to the response cue question mark). Stimulus-locked epochs lasted 150 time-frames (TF, corresponding to 293 ms) after the question mark eliciting the production of the pseudo-word; response-locked epochs lasted 150 TF (293 ms) before the vocal onset (backward) and 100 TF (195 ms) after the vocal onset (forward). The combination of the stimulus-locked and response-locked backward epochs (300 TF, 586 ms) allows to cover the mean production latencies in the slowest condition (see behavioural results). The 100 TF following the vocal onset were chosen to cover the mean duration of the onset /s/ in the sC_2C_3V condition (see Results section).

Each ERP epoch corresponding to a correct response and a valid production latency was visually inspected; epochs contaminated by eyeblinks, vertical or horizontal eye movement or other noise artefacts were rejected and excluded from averaging. After epochs inclusion, the number of averaged epochs was matched across conditions by randomly removing epochs in the conditions with more epochs. Stimulus- and response-locked ERPs were averaged separately per participant and per condition.

Electrodes presenting artefacts were interpolated using 3-D splines interpolation (Perrin et al., 1987). The average of electrodes interpolated by participants was 13.6% (max = 17%; or up to 22 out of the 128 electrodes).

ERP analyses

Analyses on the ERP data were run on waveform amplitudes and on the global distribution of the signal at scalp (microstate analyses). The first analyses were aimed at determining whether the different conditions generated different waveform amplitudes. Any difference in amplitudes can be due to different strength of the electric field or to a global topographic difference of the electric fields (revealing distinguishable brain generators), or even from latency shifts of similar brain processes. To differentiate these effects, microstate analyses based on the spatio-temporal segmentation of the signal were also performed. Microstates correspond to

stable global electrophysiological patterns at scalp – topographic maps – extending over a time-period of tenths of milliseconds, reflecting periods of stable or quasi-stable global neuronal activity; changes in the topography thus indicate changes in the global coordination of neuronal activity over time (Michel & Koenig, 2018). Given that C_1C_2V and sC_2C_3V are bound to display different amplitudes and topographies at least close to the vocal onset due to different onset phonemes (voiceless plosives in C_1C_2V and the voiceless fricative s for the sC_2C_3V stimuli, see Jouen et al., 2021), the analyses on the effect of legality was run separately for the two structures.

Waveform analysis

The ERPs were subjected to a sampling point-wise ERP waveform analysis to determine the time periods presenting local amplitude differences across legal and illegal sequences. Waveform amplitude comparisons across conditions were run using cluster-mass statistics based on permutation methods for repeated measures ANOVA (cluster mass method, 5000 permutations), with the *permuco4brain* R package (Frossard & Renaud, 2021, 2022). The analyses were run separately on stimulus-locked and on response-locked ERPs comparing two by two legal and illegal C_1C_2V and sC_2C_3V .

Microstate analyses

The aim of microstate analyses is to determine whether conditions differ in the sequences of stable or quasi-stable global electric fields (scalp topographies) (e.g. Michel et al., 2009; Michel & Murray, 2012). Changes in electric field take place when the underlying generator configuration has changed and differences in underlying generator suggest activation of different brain networks. These analyses were carried out in three steps.

At first, we run topographic consistency tests (TCT; Koenig & Melie-García, 2010) in order to verify that a given scalp field was consistently activated by the event of interest in the stimulus-locked and the response-locked data. The TCT compares, time-point by time-point, the global field power (GFP) of the averaged ERPs to the empirical distribution of the GFP obtained after the random shuffle of the data across electrodes. Stimulus – and response-locked ERPs were separately subjected to the TCT (L2 normalisation, 5000 runs and alpha of 0.05) using the Ragu software (Koenig et al., 2011).

Second, the global dissimilarity was compared across conditions. This analysis, called “TANOVA” (Murray et al., 2008) compares the global dissimilarity index (GDI, Lehmann & Skrandies, 1984) which

provides a single measure per time point reflecting the dissimilarity between two electric fields, in order to establish to what extent the ERPs' topography differs across conditions and in which time-windows. For each participant, the topographic maps obtained for each condition are re-assigned randomly (permuted) to the different experimental conditions. The group-averaged ERPs is then re-computed together with the resulting global dissimilarity index. This permutation procedure is conducted many times and the global dissimilarity index of the original ERPs is then compared to the empirical distribution resulting from the permutations. This is done to determine the likelihood of obtaining a higher global dissimilarity index value than the one actually obtained. In the present study, this analysis was conducted using the software Ragu with 5000 iterations and alpha set to 0.05.

Finally, a microstate analysis (spatio-temporal segmentation) was performed to explore whether differences in global dissimilarity obtained in the TANOVA were due to different stable topographic patterns per se or to different time course of the same stable topographic patterns across conditions. The approach used is described in Koenig et al. (2014) and implemented in the Ragu software. It determines the sequence global electrophysiological patterns at scalp (or topographic map) that best explain the data at each time frame. For the statistical analyses, the map templates identified in the spatio-temporal segmentation were "fitted" back to the individual ERPs. The fitting procedure labels each data sampling point according to the template map with which it best correlated spatially, giving as output variables the presence/duration of each map (in number of time-frames – TF – and their onset) in each individual ERP for each condition. The measures of presence/duration of each topographic map are then compared across legality conditions with non-parametrical statistics (Wilcoxon tests using the ggwithinstats R package, Patil, 2021).

Behavioural, ERP data and R-codes are available on the Yareta institutional open science archive (<https://doi.org/10.26037/yareta:yxt7iufrd5hbdi5lcz7k4qzniy>).

Results

Behavioural and acoustic results

Behavioural results are presented in Table 2. Accuracy (percentage of correct productions) was high in all conditions although lower for illegal clusters, in particular in sC₂C₃V stimuli.

Initialisation times were slightly faster for illegal than for legal C₁C₂V and sC₂C₃V (see Table 2). This was

Table 2. Production latencies, acoustic duration and accuracy for legal and illegal C₁C₂V and sC₂C₃V stimuli (standard deviation (SD) in parentheses).

	Production latencies (SD)	Acoustic duration (SD)	Accuracy (SD)
C ₁ C ₂ V			
legal	600.2 (97.0)	326.8 (33.4)	96.3% (3.1)
illegal	588.6 (90.6)	372.9 (33.5)	91.8% (7.7)
sC ₂ C ₃ V			
legal	526.8 (87.3)	496.8 (52.6)	97.0% (3.8)
illegal	522.1 (87.7)	541.1 (42.7)	86.6% (8.4)

confirmed by the statistical model ($[ProductionLatency \square Legality * Structure + Order + (1|participant) + (1|stimuli)]$), which revealed a main effect of legality ($F(1, 109) = 4.32, p = .039$). There was also a main effect of structure ($F(1, 109) = 273.29, p < .001$), with sC₂C₃V stimuli being initialised faster than C₁C₂V and no interaction between the two factors ($F < 1$).

The acoustic duration was bimodally distributed when the two structures were collapsed, due to longer duration for sC₂C₃V, and normally distributed for each structure separately. The analyses were run separately for C₁C₂V and sC₂C₃V. In both structures, the statistical models ($[Duration \square Legality + Order + (1|participant) + (1|stimuli)]$), indicate that illegal stimuli were about 50 ms longer than the corresponding legal stimuli (C₁C₂V: $t(53) = -18.65, p < .0001, SE = 3.05, Estimate = -56.97$; sC₂C₃V: $t(52) = -12.85, p < .0001, SE = 3.54, Estimate = -45.53$).

The duration of /s/ in the sC₂C₃V stimuli was 197 and 195 ms respectively for legal and illegal stimuli, with no significant difference ($F < 1$).

ERP results

The results for the stimulus-locked ERPs confirm that legality conditions do not differ in the time-window locked to the question mark eliciting the response neither in the waveform analyses nor in the microstate analyses and are presented in the Appendix B. Only the results on the response-locked ERP signal are detailed below.

Waveform analysis

The results of the waveform amplitude analyses are presented in Figure 1(a). In the response-locked ERPs, different amplitudes between legal and illegal sequences were observed only for the C₁C₂V stimuli in the entire time window preceding the vocal onset (Figure 1(a), left panel). Amplitudes differed on a large cluster (28 neighbouring electrodes) of anterior

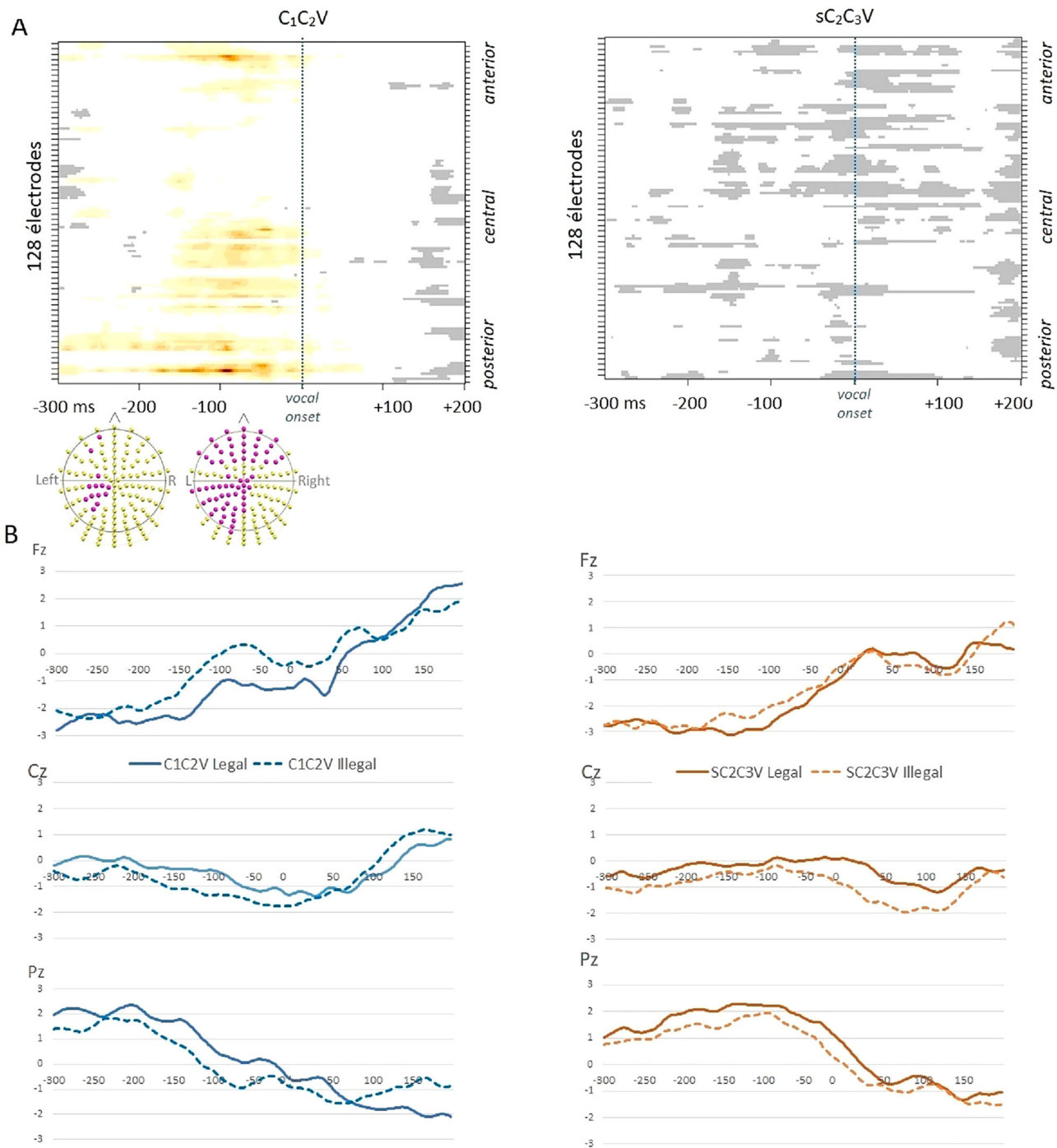


Figure 1. Results of the ERP waveform analyses on response-locked data on legal and illegal C_1C_2V on the left hand-side and legal and illegal sC_2C_3V on the right. (a) Results of the cluster mass waveform analyses, with periods of significant p values (in red $p < 0.01$ and in yellow $p < 0.05$, in white and grey, $p \geq 0.05$) on electrodes (Y axes) and time points (X axes), with electrodes yielding significant differences in amplitudes around -200 ms and around -100 ms highlighted in pink on the arrangement of the 128 electrodes. (b) Exemplar waveforms on Cz, Fz and Pz for each condition.

right and left electrodes and on a cluster of about 30 central-posterior electrodes in the last -150 ms preceding the vocal onset, and on a smaller cluster of central-left electrodes (around 10 neighbouring electrodes) in the -150 to -300 ms preceding the vocal onset. In these time-windows, amplitudes are smaller

or less negative for illegal than for legal C_1C_2V stimuli on anterior electrodes (see Fz in Figure 1(b)) and less positive on central and posterior channels (see Cz and Pz). No significant effect of legality was found on amplitudes for the sC_2C_3V stimuli (right panel of Figure 1(a)).

TANOVA

Pairwise TANOVAs revealed significant differences in the comparisons between legal and illegal C_1C_2V response-locked ERPs from ~ -150 ms to the vocal onset (0 ms).

These results indicate that the global electric fields differ between legal and illegal C_1C_2V in the same -150 ms preceding the vocal onset where different amplitudes are also observed. On the sC_2C_3V the

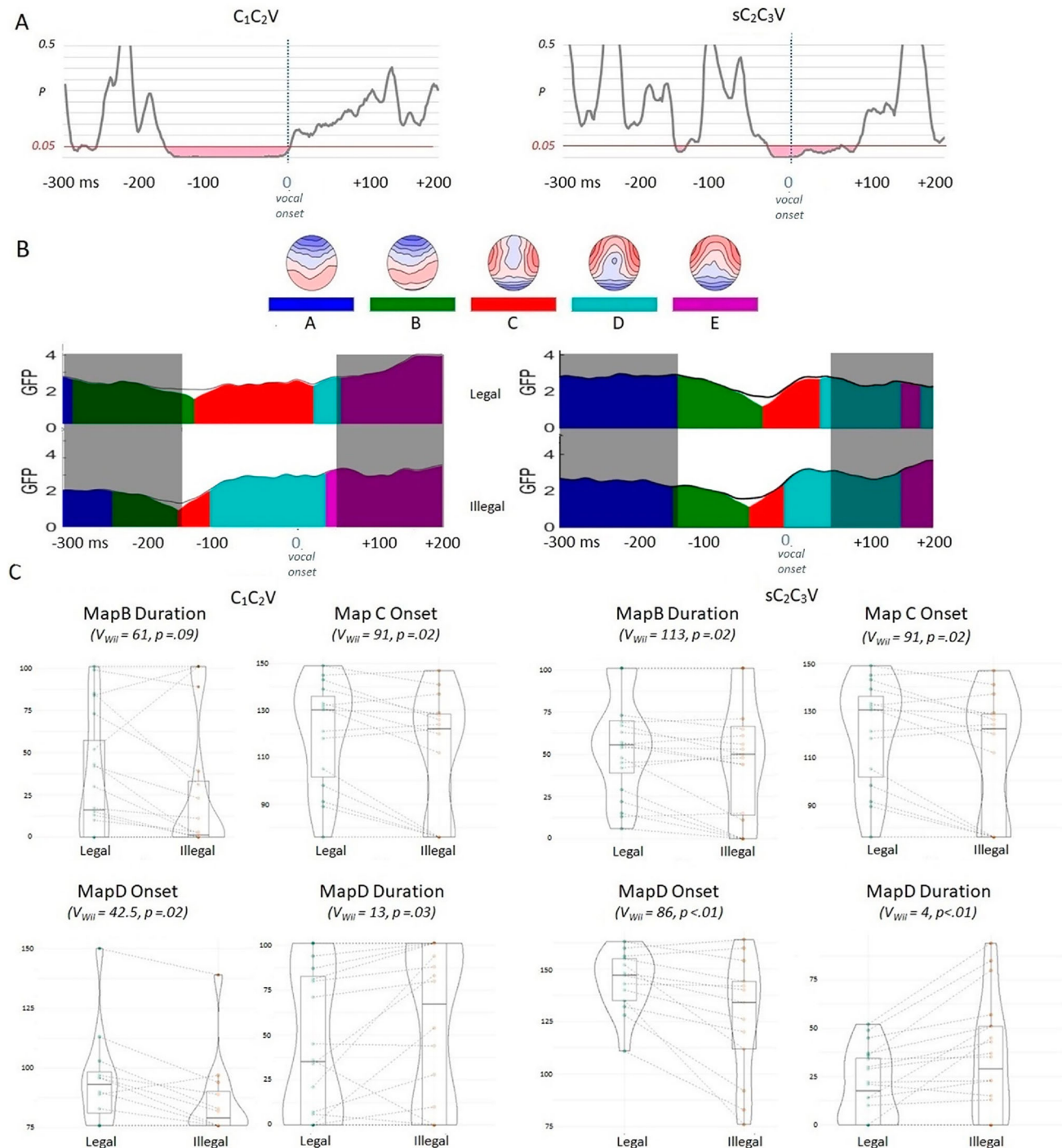


Figure 2. Results of the ERP microstate (topographic) analyses on response-locked data, with legal and illegal C_1C_2V on the left hand-side and legal and illegal sC_2C_3V on the right. (a) Results of the TANOVA, with the red horizontal bar representing $\alpha = 0.05$ and periods of significant p -values coloured in pink. (b) Results of the spatio-temporal segmentation, with different colours under the GFP representing periods of stability of the electrophysiological pattern at scalp, related to the five corresponding map templates (A, B, C, D and E) displayed on the first line. The shaded periods correspond to the time-windows excluded from the fitting procedure. (c) Violin plots and statistical results for the fitting in the individual ERP of the legal and illegal conditions.

production of legal and illegal sequences differed on global electric field shortly (30 ms) before the vocal onset and during the first 75 ms following the vocal onset (see Figure 2(a)). These differences in global electric field may reflect either different microstates, corresponding to the activation of different brain networks, or different duration and/or temporal shifts of the same activations in these time-windows. The results of the microstate analyses will further clarify on these possible interpretations of the TANOVA results.

Microstate analyses

The spatio-temporal segmentation of the entire response-locked ERPs revealed five different microstates (labelled “A” to “E” in Figure 2(b)) accounting for 97.3% of the variance. The same microstates are observed in all conditions, but with different duration and consequently different time-distribution, A different distribution of the microstates preceding the vocal onset was expected for C_1C_2V and sC_2C_3V sequences because of different properties of the onset phonemes (plosives versus fricatives, see method section) and will not be analysed further here; only within C_1C_2V and within sC_2C_3V analyses on legality will be run.

The fitting of map templates in the individual ERPs was done in the time window yielding significant TANOVA across legality conditions, namely from 75 to 175 TF (corresponding to -150 to $+100$ ms relative to the vocal onset), with the microstate maps B, C and D. The results of the fitting statistics are presented in Figure 2(c).

No significant difference ($p = .09$) across legal and illegal conditions is observed on map B for the C_1C_2V stimuli while it is significantly more present in the legal than illegal sC_2C_3V sequences (respectively 55.9 and 48.6 TF). For both structures, the onset of the following map C is delayed in the legal condition (mean onset of map C: -45.6 F (-89 ms) before the vocal onset for legal C_1C_2V versus -57.3 TF (-112 ms) for illegal C_1C_2V ; -30.6 TF (-59.8 ms) before the vocal onset for legal sC_2C_3V versus -38.9 TF (-76 ms) for illegal sC_2C_3V). No further significant differences across conditions appear on Map C, but the onset of the following microstate Map D is also significantly delayed in the legal stimuli for both structures. Map D is also significantly more present in the illegal stimuli in both structures (mean duration of map D in TF in C_1C_2V : 43.25 versus 54.35; in sC_2C_3V : 19.6 versus 32.6).

In sum, the results of the microstate analysis indicate that the same microstates (i.e. the activation of the same brain networks) are observed in all conditions, but with different activation in time (a later onset of Maps C and D

for legal sequences). This temporal shift is similar for C_1C_2V and sC_2C_3V .

Discussion

In the present study, we contrasted the delayed production of closely matched legal and illegal CCV and sCCV sequences. The illegal sequences are composed of consonants that can occur in succession (e.g. across successive words) but that do not form a possible syllable onset in French. Consequently, they are unlikely stored and retrieved as a whole syllable, it is even questionable whether they are assembled as a single speech plan. Interestingly, the behavioural results were opposite to the results of previous studies on production latencies, as illegal sequences were initialised slightly faster than legal syllables independently of the sequence structure (C_1C_2 or sC_2C_3). The acoustic durations of the productions were longer for illegal than for legal sequences. Finally, ERP results indicated different waveform amplitudes between legal and illegal sequences in the response-locked signal only for C_1C_2V sequences as well as a pattern of different time-distributions of the same microstates between legal and illegal sequences, pattern that is similar for C_1C_2V and sC_2C_3V .

In the following we will discuss the behavioural, acoustic and ERP results separately first, before an interpretation based on the integration of all results.

Shorter latencies for illegal sequences

At first sight the faster initialisation of illegal sequences is counterintuitive and in contradiction with the results of previous studies (showing slower initialisation for infrequent syllables as compared to frequent syllables, see Introduction). It is worth reminding that the range of the effect in studies reporting faster initialisation speed for frequent relative to infrequent syllables in delayed production tasks is within the range of the difference observed here according to legality, although in the opposite direction (5 ms for monosyllabic and 14 ms for disyllabic stimuli in Cholin et al., 2011; 20 ms in Lagano & Alario, 2006; 20 ms in Bürki et al., 2015).

A closer look to previous studies also indicates that faster initialisation for speech sequences containing illegal sequences has already been observed. Indeed, Segawa et al. (2015), who used a similar paradigm (a delayed production task with legal and illegal German sequences in monosyllabic pseudowords), reported 40 ms faster initialisation for illegal sequences, but without providing an interpretation for this result. In the present study, the faster initiation of illegal syllables

cannot be due to the acoustic properties of the onset phoneme as they are perfectly matched across legal and illegal stimuli. In addition, the observation of the same effect in the CCV and sCCV structures, along with the convergence with previous results by Segawa et al. (2015), makes these counterintuitive results reliable and worth a discussion.

The reasons why such results have not been observed in other studies using (standard) delayed production tasks may be related to the type of stimuli or to differences in the paradigms. Indeed, as already mentioned in the Introduction, most previous studies contrasted legal syllables of high and low frequency and did not report different initialisation latencies when (standard) delayed production paradigms were used (see Laganaro & Alario, 2006). As for the studies by Bürki et al. (2015, 2020) who did contrast frequent legal syllable to novel illegal sequences in delayed production, the stimuli were disyllabic pseudo-words and, more importantly, the paradigm (a delayed production in a phoneme completion task) was different from the standard delayed production used in the present study and in Segawa et al. (2015).

The present result and that of Segawa et al. (2015) converge in showing that some encoding processes are computed faster – and at least not slower – for illegal syllables than for legal ones when the production can be prepared and is triggered by a cue. In the framework of models presented in the Introduction, the production of speech sequences that are uncommon/untrained has been claimed to involve the assemblage of elements into a speech plan, process that is claimed to induce a larger computation cost relative to the retrieval of routinised plans from memory. It has also been suggested that speech plans can be prepared during the delay (if it is not filled with articulatory suppression, Laganaro & Alario, 2006). Following this rationale, our results do suggest a reduced cost for illegal sequences in the processes that span from phonetic planning to articulation. These processes correspond to the motor programming level in models acknowledging multiple speech preparation processes (Guenther, 2016; Van Der Merwe, 2021, see Introduction section). A reduced motor programming time for illegal sequences could be interpreted in two different ways. First, articulatory studies have shown that illegal sequences of consonants are timed differently, with less or more variable overlap relative to legal syllable onsets (Davidson, 2006). For these last ones, a tight coordination is required within the consonants and between the onset cluster and the vowel nucleus (Marin & Pouplier, 2010). It is thus possible that the muscle-specific motor programmes are computed faster when the speech elements comprised in

the speech plan are loosely timed (for illegal sequences) than when they are tightly timed (for legal syllables). Alternatively, the motor preparation may be faster for illegal sequences because it operates on smaller speech plans than for legal syllables. If the illegal sequence is encoded in several speech plans (e.g. $C_1 + C_2V$), then the production of the sequence can be initialised as soon as the muscle-specific programmes are encoded for the first available unit (e.g. C_1), while the entire syllable-size speech plan (e.g. C_1C_2V) needs to be parameterised before articulation can start for the legal syllables. Here the initialisation difference between legal and illegal sequences seems larger for C_1C_2V than for sC_2C_3V , which may be related to the position of the illegal CC cluster (C_1C_2 versus C_2C_3), or to specific properties of the initial /s/, which will be further discussed in the final section. In any case, disentangling those two interpretations is not possible based on behavioural results alone, and we will come back to it after the discussion of the duration and ERP results.

Longer acoustic durations for illegal syllables

The illegal sequences are found to be about 50 ms longer than their respective legal counterparts, independent of the structure of the sequence (CCV or sCCV). This difference can not be attributed to the segmental composition of the sequence since the illegal sequences differ from legal syllables by the substitution of a single phoneme (replacing the liquid /R/ or /l/ with /n/, /m/, /p/, /t/, /v/ or /z/) which duration is similar to the liquid consonants (O'Shaughnessy, 1981). The longer duration of the illegal sequences should rather be attributed to the fact that consonants which do not form a legal syllable onset are sequentially organised (i.e. are less temporally overlapped), as shown in other studies (Bombien & Hoole, 2013; Davidson & Stone, 2003; Davidson, 2006; Kühnert et al., 2006). As for languages where consonant clusters are not allowed as syllable onsets (e.g. varieties of Arabic, Gafos et al., 2020), it is probable that in our illegal sequences only the immediately prevocalic consonant enters into a stable relation with the vowel nucleus, while the other preceding consonants are planned to be initiated sequentially.

Different ERP correlates preceding the vocal onset (in C_1C_2V stimuli)

The time-window of ~150 ms preceding the vocal onset of legal and illegal CCV syllables differed on waveform amplitudes and on the stable global field topographies. While the stimulus-locked ERPs likely reflect mental processes related to the processing of the visual cue

triggering the delayed production (see for instance the map template “2” in Appendix B corresponding to the classic P1 VEP component), the response-locked ERPs are thought to reflect speech encoding, here during the 300 ms preceding the vocal onset. In the time-window of ERP differences between legal and illegal C_1C_2V , two phenomena are observed. First, the onset of map C is delayed of about 22 ms in the legal relative to the illegal syllables; second, the onset of the following microstate Map D is also significantly delayed in the legal stimuli and Map D also lasts about 25 ms longer in the production of the illegal stimuli. The topography of this latter microstate (map D) has previously been associated with articulatory movement (Jouen et al., 2021): its early onset, relative to the vocal onset, probably corresponds to articulatory movements preceding the vocal signal. Indeed, Map D starts much later for the stimuli beginning with /s/ than for stimuli beginning with an unvoiced stop, in-line with longer articulatory to acoustic lag for unvoiced stops corresponding to the silent closure (Mooshammer et al., 2012; Rastle et al., 2005).

The topographic map C is present in all conditions, but over a short time-window in most conditions. Based on its brief presence in three conditions and on its topography, it could be interpreted as a transition between map B and D. However, such interpretation only fits with its short duration in the sC_2C_3V and illegal C_1C_2V , not with the large time-window covered by map C in the legal C_1C_2V . The differences in waveform amplitudes, in the TANOVA and in the fitting results in the 150 ms preceding the vocal onset for the C_1C_2V stimuli (along with similar results for the sC_2C_3V stimuli, see next sections) point to different timing for brain processes that are engaged in the parametrisation of legal relative to illegal syllable onsets. In other words, the microstates do not differ across conditions, but are distributed differently. Along with larger amplitudes in the legal condition (see Figure 1(b)), a different intensity/dynamics of recruitment of the same brain areas for legal and illegal C_1C_2V seems the most plausible explanation, – as further discussed in the last section–.

Different ERP correlates around the vocal onset (in sC_2C_3V stimuli)

For the sC_2C_3V stimuli, no significant differences in waveform amplitudes were observed between legal and illegal syllables but the TANOVA indicated differences in the global electric fields in a time window of about 100 ms around the vocal onset. Here again, the microstates do not differ across legality conditions but are differently distributed, with a longer lasting Map B

in the legal sequences and delayed onsets of maps C and D. It is worth reminding that the ERP differences between legal and illegal sequences observed in the first 75 ms following the vocal onset in sC_2C_3V stimuli fall in the time-window of articulation of /s/, that is same across legal and illegal conditions along with C_2 . The delayed onset of Maps C and D for legal sequences and longer lasting map D in the production of illegal sequences is entirely coherent with the results of the C_1C_2V syllables, except that the same microstates occur later (owing to the shift of maps C and D due to different onset consonants in C_1C_2V versus sC_2C_3V , as already discussed in the previous section). We can thus consider that the results on microstates partly mirror those described for C_1C_2V and will therefore be discussed together in the following section.

Integration of behavioural, acoustic and ERP results

Previous studies have sought larger processing costs for infrequent or illegal syllables, and have interpreted it in the framework of models suggesting retrieval of syllable-sized speech plans for frequent speech sequences versus assembled plans from smaller units for infrequent or illegal sequences (see Introduction). The current results cannot be interpreted along these same lines for two reasons. First, the present results are obtained with a standard delayed production task, in which the speakers can prepare the phonetic plans during the delay and different EEG/ERP patterns are observed immediately preceding and during speech articulation. Second, the behavioural results are in the opposite direction (larger preparation costs in terms of production latencies for legal than for illegal sequences). As anticipated in the discussion of the behavioural results above, the ERP results indicate that some processes related to motor programming differ (in terms of cost and dynamics of mental processes) between legal and illegal sequences. Among the possible interpretations put forward for the behavioural results, the suggestion that motor speech programming operates on speech plans of different sizes for legal and illegal syllables seems to be compatible also with the ERP results that indicate differently distributed/involved brain processes. In particular, the shorter initialisation times for illegal sequences and the earlier onsets of several periods of stable electrophysiological activity at scalp (microstates C and D) are compatible with the idea that the production of illegal sequences is initialised as soon as the muscle-specific programmes are encoded for the first available unit (C_1); conversely, longer vocal onset time and later

onsets of microstates C and D for the legal syllables are coherent with a larger size of prepared units. Legal consonant clusters at the onset of syllables are organised as a whole unit with respect to the following vowel. This organisation relies on a competitive coupling relation with, on one side, all the consonant gestures involved coupled in-phase to the vowel nucleus and, on the other side, the consonants coupled antiphase with each other (Marin & Pouplier, 2010), i.e. they are planned as a time-organised ensemble of articulatory gestures. In case of illegal CC, the gestures are probably not timed together and/or are planned to be initiated serially/sequentially. It means that, for illegal syllables, articulation may start as soon as the first consonant is ready and then, the parametrisation of temporal organisation between the elements of the sequence would be done on the fly, which is also compatible with the longer acoustic duration.

This interpretation perfectly matches the different results for the C_1C_2V sequences, for which all differences between legal and illegal sequences are observed in the 150 ms preceding the vocal onset. By contrast the pattern of ERP differences between legal and illegal sequences seems to be more limited for sC_2C_3V syllables. The later onset of the microstates associated with articulatory movements for the sC_2C_3V syllables has been discussed in the previous section, it does however not account for the absence of any difference before the vocal onset for the production of legal versus illegal sC_2C_3V sequences. Why do legal and illegal sC_2C_3V elicit different electrophysiological signatures only after the start of the /s/ articulation? The syllabification of $sC(C)$ sequences at the beginning of words is language specific and it has been claimed to be extrasyllabic in French (Rialland, 1994), as suggested also for other languages such as Italian based on articulatory investigation of consonant timing (Hermes et al., 2013). Under this assumption, the temporal orchestration of the gestures for the legal $sCCV$ syllable in French should mirror that found in Italian: a cohesive CCV syllabic organisation but a sequential organisation for the preceding /s/. In the case of the legal $sCCV$ sequences, we can interpret the fact that both the legal and the illegal $sCCV$ sequences are encoded in several sequentially organised units ($s + CCV$ for the legal syllables and $s + C + CV$ for illegal sequences) and therefore, they do not trigger clearly different ERP patterns. Only once the /s/ is fully programmed and ready to be articulated, different processes seem to underlie the preparation of the following legal versus illegal C_2C_3 sequence.

However, we are conscious that the acoustic information taken as vocal onset does not directly correspond to the onset of articulatory gestures. Indeed the

acoustic is blind for the onset of closure formation for the voiceless stops, for the muscular initiation of articulatory and respiratory movements, for speech ready gestures, etc ... Further investigations of EEG/ERP activity with a direct observation of articulation is experimentally challenging, but is necessary to validate further these interpretations.

Conclusions

In previous studies, differences in behaviour and in brain activation for the production of highly practiced (frequent) vs. uncommon (infrequent or illegal) speech sequences have been attributed to the retrieval versus assembling of phonetic plans. Here we showed that the encoding of practiced (legal) and uncommon (illegal) syllables differs also beyond phonetic planning, and that illegal syllables are ready for execution faster than the legal syllables. Our acoustic and ERP results are compatible with different mechanisms and/or different sizes of motor programming/execution for legal and illegal sequences: the units for the illegal sequences are sent sequentially to execution, while the execution of the syllable-size plan for the legal syllables (although potentially retrieved as a whole from syllabary) can start only after the parametrisation of a whole sequence.

Further investigations are necessary to complete the understanding of motor speech encoding of legal versus illegal sequences at phonetic encoding and beyond and of the size of speech sequences sent to execution. Future studies may take advantage of contrasting different conditions of delayed production, while analysing brain activation in a time-window immediately preceding and following the vocal onset. Additionally, some developments are still necessary to tease apart effects related to motor speech encoding processes from those related to articulation.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by Swiss National Science Foundation under Grants CRSII5_173711 and CRSII5_202228.

Data availability statement

The data used in the study will be made available on the Yareta institutional open science archive (<https://doi.org/10.26037/yareta:yxt7iufrd5hbdi5lcz7k4qznyi>).

References

- Aichert, I., & Ziegler, W. (2004). Syllable frequency and syllable structure in apraxia of speech. *Brain and Language*, 88(1), 148–159. [https://doi.org/10.1016/S0093-934X\(03\)00296-7](https://doi.org/10.1016/S0093-934X(03)00296-7)
- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59(4), 390–412. <https://doi.org/10.1016/j.jml.2007.12.005>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1). <https://doi.org/10.18637/jss.v067.i01>
- Bohland, J. W., & Guenther, F. H. (2006). An fMRI investigation of syllable sequence production. *NeuroImage*, 32(2), 821–841. <https://doi.org/10.1016/j.neuroimage.2006.04.173>
- Bombien, L., & Hoole, P. (2013). Articulatory overlap as a function of voicing in French and German consonant clusters. *The Journal of the Acoustical Society of America*, 134(1), 539–550. <https://doi.org/10.1121/1.4807510>
- Brendel, B., Erb, M., Riecker, A., Grodd, W., Ackermann, H., & Ziegler, W. (2011). Do we have a “mental syllabary” in the brain? An fMRI study. *Motor Control*, 15(1), 34–51. <https://doi.org/10.1123/mcj.15.1.34>
- Brunet, D., Murray, M. M., & Michel, C. M. (2011). Spatiotemporal analysis of multichannel EEG: CARTOOL. *Computational Intelligence and Neuroscience*, 2011, 1–15.
- Bürki, A., Cheneval, P. P., & Laganaro, M. (2015). Do speakers have access to a mental syllabary? ERP comparison of high frequency and novel syllable production. *Brain and Language*, 150, 90–102.
- Bürki, A., Viebahn, M., & Gafos, A. (2020). Plasticity and transfer in the sound system: Exposure to syllables in production or perception changes their subsequent production. *Language, Cognition and Neuroscience*, 35(10), 1371–1393.
- Carreiras, M., Mechelli, A., & Price, C. J. (2006). Effect of word and syllable frequency on activation during lexical decision and reading aloud. *Human Brain Mapping*, 27(12), 963–972. <https://doi.org/10.1002/hbm.20236>
- Chang, S.-E., Kenney, M. K., Loucks, T. M. J., Poletto, C. J., & Ludlow, C. L. (2009). Common neural substrates support speech and non-speech vocal tract gestures. *NeuroImage*, 47(1), 314–325. <https://doi.org/10.1016/j.neuroimage.2009.03.032>
- Cholin, J., Dell, G. S., & Levelt, W. J. M. (2011). Planning and articulation in incremental word production: Syllable-frequency effects in English. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37(1), 109–122. <https://doi.org/10.1037/a0021322>
- Cholin, J., Levelt, W., & Schiller, N. (2006). Effects of syllable frequency in speech production. *Cognition*, 99(2), 205–235. <https://doi.org/10.1016/j.cognition.2005.01.009>
- Code, C. (1998). Models, theories and heuristics in apraxia of speech. *Clinical Linguistics & Phonetics*, 12(1), 47–65. <https://doi.org/10.3109/02699209808985212>
- Darley, F. L., Aronson, A. E., & Brown, J. R. (1975). *Motor speech disorders*. WB Saunders Company.
- Davidson, L. (2006). Phonology, phonetics, or frequency: Influences on the production of non-native sequences. *Journal of Phonetics*, 34(1), 104–137. <https://doi.org/10.1016/j.wocn.2005.03.004>
- Davidson, L., & Stone, M. (2003). Epenthesis versus gestural mistiming in consonant cluster production: An ultrasound study. In *Proceeding of the West Coast Conference on Formal Linguistics* (Vol. 22, pp. 165–178).
- Duffy, J. R. (2019). *Motor speech disorders e-book: Substrates, differential diagnosis, and management*. Elsevier Health Sciences.
- Frossard, J., & Renaud, O. (2021). Permutation tests for regression, ANOVA, and comparison of signals: The Permuc package. *Journal of Statistical Software*, 99(15), 1–32. <https://doi.org/10.18637/jss.v099.i15>
- Frossard, J., & Renaud, O. (2022). The cluster depth tests: Toward point-wise strong control of the family-wise error rate in massively univariate tests with application to M/EEG. *NeuroImage*, 247, 118824. <https://doi.org/10.1016/j.neuroimage.2021.118824>
- Gafos, A. I., Hoole, P., Roon, K., & Zeroual, C. (2020). Variation in overlap and phonological grammar in Moroccan Arabic clusters. In C. Fougeron, B. Kühnert, M. d’Imperio, & N. Vallée (Eds.), *Laboratory phonology* (Vol. 10, pp. 657–698). Mouton de Gruyter.
- Guenther, F. H. (2016). *Neural control of speech*. MIT Press.
- Guenther, F. H., Ghosh, S. S., & Tourville, J. A. (2006). Neural modeling and imaging of the cortical interactions underlying syllable production. *Brain and Language*, 96(3), 280–301. <https://doi.org/10.1016/j.bandl.2005.06.001>
- Hermes, A., Mücke, D., & Grice, M. (2013). Gestural coordination of Italian word-initial clusters: The case of ‘impure s’. *Phonology*, 30(1), 1–25.
- Jouen, A.-L., Lancheros, M., & Laganaro, M. (2021). Microstate ERP analyses to pinpoint the articulatory onset in speech production. *Brain Topography*, 34(1), 29–40. <https://doi.org/10.1007/s10548-020-00803-3>
- Koenig, T., Kottlow, M., Stein, M., & Melie-García, L. (2011). Ragu: A free tool for the analysis of EEG and MEG event-related scalp field data using global randomization statistics. *Computational Intelligence and Neuroscience*, 2011, 1–14.
- Koenig, T., & Melie-García, L. (2010). A method to determine the presence of averaged event-related fields using randomization tests. *Brain Topography*, 23(3), 233–242. <https://doi.org/10.1007/s10548-010-0142-1>
- Koenig, T., Stein, M., Grieder, M., & Kottlow, M. (2014). A tutorial on data-driven methods for statistically assessing ERP topographies. *Brain Topography*, 27, 72–83.
- Krakow, R. A. (1999). Physiological organization of syllables: A review. *Journal of Phonetics*, 27(1), 23–54. <https://doi.org/10.1006/jpho.1999.0089>
- Kröger, B. J., Kannampuzha, J., & Neuschaefer-Rube, C. (2009). Towards a neurocomputational model of speech production and perception. *Speech Communication*, 51(9), 793–809. <https://doi.org/10.1016/j.specom.2008.08.002>
- Kühnert, B., Hoole, P., & Mooshammer, C. (2006). Gestural overlap and C-center in selected French consonant clusters. In *7th International Seminar on Speech Production (ISSP)* (p. 327).
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82(13), 1–26. <https://doi.org/10.18637/jss.v082.i13>
- Laganaro, M. (2005). Syllable frequency effect in speech production: Evidence from aphasia. *Journal of Neurolinguistics*,

- 18(3), 221–235. <https://doi.org/10.1016/j.jneuroling.2004.12.001>
- Laganaro, M., & Alario, F.-X. (2006). On the locus of the syllable frequency effect in speech production. *Journal of Memory and Language*, 55(2), 178–196. <https://doi.org/10.1016/j.jml.2006.05.001>
- Laganaro, M., Croisier, M., Bagou, O., & Assal, F. (2012). Progressive apraxia of speech as a window into the study of speech planning processes. *Cortex*, 48(8), 963–971.
- Lancheros, M., Jouen, A., & Laganaro, M. (2020). Neural dynamics of speech and non-speech motor planning. *Brain and Language*, 203, 104742. <https://doi.org/10.1016/j.bandl.2020.104742>
- Lehmann, D., & Skrandies, W. (1984). Spatial analysis of evoked potentials in man—A review. *Progress in Neurobiology*, 23(3), 227–250.
- Levelt, W. J. M. (1995). *Speaking: From intention to articulation* (4th ed.). The MIT Press.
- Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, 22(1), 1–38. <https://doi.org/10.1017/S0140525X99001776>
- Levelt, W. J. M., & Wheeldon, L. (1994). Do speakers have access to a mental syllabary? *Cognition*, 50(1–3), 239–269. [https://doi.org/10.1016/0010-0277\(94\)90030-2](https://doi.org/10.1016/0010-0277(94)90030-2)
- Marin, S., & Pouplier, M. (2010). Temporal organization of complex onsets and codas in American English: Testing the predictions of a gestural coupling model. *Motor Control*, 14(3), 380–407. <https://doi.org/10.1123/mcj.14.3.380>
- Michel, C. M., & Koenig, T. (2018). EEG microstates as a tool for studying the temporal dynamics of whole-brain neuronal networks: A review. *Neuroimage*, 180, 577–593. <https://doi.org/10.1016/j.neuroimage.2017.11.062>
- Michel, C. M., Koenig, T., Brandeis, D., & Gianotti, L. R. R. (2009). *Electric neuroimaging*. Cambridge University Press.
- Michel, C. M., & Murray, M. M. (2012). Towards the utilization of EEG as a brain imaging tool. *Neuroimage*, 61(2), 371–385. <https://doi.org/10.1016/j.neuroimage.2011.12.039>
- Mooshammer, C., Goldstein, L., Nam, H., McClure, S., Saltzman, E., & Tiede, M. (2012). Bridging planning and execution: Temporal planning of syllables. *Journal of Phonetics*, 40(3), 374–389. <https://doi.org/10.1016/j.wocn.2012.02.002>
- Moser, D., Fridriksson, J., Bonilha, L., Healy, E. W., Baylis, G., Baker, J. M., & Rorden, C. (2009). Neural recruitment for the production of native and novel speech sounds. *Neuroimage*, 46(2), 549–557. <https://doi.org/10.1016/j.neuroimage.2009.01.015>
- Murray, M. M., Brunet, D., & Michel, C. (2008). Topographic ERP analyses: A step-by-step tutorial review. *Brain Topography*, 20(4), 249–264. <https://doi.org/10.1007/s10548-008-0054-5>
- New, B., Pallier, C., Brysbaert, M., & Ferrand, L. (2004). Lexique 2: A new French lexical database. *Behavior Research Methods, Instruments, and Computers*, 36(3), 516–524.
- O’Shaughnessy, D. (1981). A study of French vowel and consonant durations. *Journal of Phonetics*, 9(4), 385–406. [https://doi.org/10.1016/S0095-4470\(19\)31016-2](https://doi.org/10.1016/S0095-4470(19)31016-2)
- Papoutsis, M., de Zwart, J. A., Jansma, J. M., Pickering, M. J., Bednar, J. A., & Horwitz, B. (2009). From phonemes to articulatory codes: An fMRI study of the role of Broca’s area in speech production. *Cerebral Cortex*, 19(9), 2156–2165. <https://doi.org/10.1093/cercor/bhn239>
- Parrell, B., Lammert, A. C., Ciccarelli, G., & Quatieri, T. F. (2019). Current models of speech motor control: A control-theoretic overview of architectures and properties. *The Journal of the Acoustical Society of America*, 145(3), 1456–1481.
- Patil, I. (2021). Visualizations with statistical details: The ‘ggstatsplot’ approach. *Journal of Open Source Software*, 6(61), 3167. <https://doi.org/10.21105/joss.03167>
- Perrin, F., Pernier, J., Bertrand, O., Giard, M.-H., & Echallier, J. (1987). Mapping of scalp potentials by surface spline interpolation. *Electroencephalography and Clinical Neurophysiology*, 66(1), 75–81. [https://doi.org/10.1016/0013-4694\(87\)90141-6](https://doi.org/10.1016/0013-4694(87)90141-6)
- Protopapas, A. (2007). Check Vocal: A program to facilitate checking the accuracy and response time of vocal responses from DMDX. *Behavior Research Methods*, 39(4), 859–862. <https://doi.org/10.3758/BF03192979>
- Rastle, K., Croot, K. P., Harrington, J. M., & Coltheart, M. (2005). Characterizing the motor execution stage of speech production: Consonantal effects on delayed naming latency and onset duration. *Journal of Experimental Psychology: Human Perception and Performance*, 31(5), 1083–1095. <https://doi.org/10.1037/0096-1523.31.5.1083>
- Rialland, A. (1994). The phonology and phonetics of extrasyllabicity in French. In P. A. Keating (Ed.), *Phonological structure and phonetic form* (Vol. 3, pp. 136–159). Cambridge University Press.
- Schiller, N. O., Meyer, A. S., Baayen, R. H., & Levelt, W. J. M. (1996). A comparison of lexeme and speech syllables in Dutch. *Journal of Quantitative Linguistics*, 3(1), 8–28. <https://doi.org/10.1080/09296179608590060>
- Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime reference guide*. Psychology Software Tools, Incorporated.
- Segawa, J. A., Tourville, J. A., Beal, D. S., & Guenther, F. H. (2015). The neural correlates of speech motor sequence learning. *Journal of Cognitive Neuroscience*, 27(4), 819–831. https://doi.org/10.1162/jocn_a_00737
- Staiger, A., & Ziegler, W. (2008). Syllable frequency and syllable structure in the spontaneous speech production of patients with apraxia of speech. *Aphasiology*, 22(11), 1201–1215. <https://doi.org/10.1080/02687030701820584>
- Tilsen, S., Spincemaille, P., Xu, B., Doerschuk, P., Luh, W.-M., Feldman, E., & Wang, Y. (2016). Anticipatory posturing of the vocal tract reveals dissociation of speech movement plans from linguistic units. *PLoS one*, 11(1), e0146813. <https://doi.org/10.1371/journal.pone.0146813>
- Van Der Merwe, A. (2021). New perspectives on speech motor planning and programming in the context of the four-level model and its implications for understanding the pathophysiology underlying apraxia of speech and other motor speech disorders. *Aphasiology*, 35(4), 397–423. <https://doi.org/10.1080/02687038.2020.1765306>
- Ziegler, W. (2009). Modelling the architecture of phonetic plans: Evidence from apraxia of speech. *Language and Cognitive Processes*, 24(5), 631–661. <https://doi.org/10.1080/01690960802327989>