Ocular dynamics reveal articulatory processing at single-phoneme level during silent reading

Alan Taitz*[1], Diego E Shalom[2] and Marcos A Trevisan[1,2]

[1] Physics Institute of Buenos Aires (IFIBA) CONICET, Buenos Aires, Argentina;
[2]Department of Physics, University of Buenos Aires (UBA), Buenos Aires 1428EGA, Argentina.

*taitz@df.uba.ar

Silent reading is a cognitive operation that produces verbal content with no vocal output. One relevant question is the extent to which this verbal content is processed as overt speech in the brain. To address this, we investigated the signatures of articulatory processing during reading. We acquired sound, eye trajectories and vocal gestures during the reading of consonant-consonant-vowel (CCV) pseudowords. We found that the duration of the first fixations on the CCVs during silent reading are correlated to the duration of the transitions between consonants when the CCVs are actually uttered. An articulatory model of the vocal system was implemented to show that consonantal transitions measure the articulatory effort required to produce the CCVs. These results demonstrate that silent reading is modulated by slight articulatory features such as the laryngeal abduction needed to devoice a single consonant or the reshaping of the vocal tract between successive consonants.

silent reading | articulatory processing | vocal models

Introduction

The faculty of language entails a repertoire of mental operations known as inner speech, in which verbal content is produced but voice is inhibited [1]. This internal production of words is ubiquitous: “as you read this text, the chances are you can hear your own inner voice narrating the words. You may hear your inner voice again when […] imagining how a phone conversation this afternoon will play out” [2].

One relevant question regarding speech-related tasks is the extent to which they are processed as actual speech in the brain. During speech perception, for example, neural patterns are organized around acoustic features and do not contain articulatory representations as the ones produced during speech [3]. On the contrary, speech imagery has been recently associated with the production of efference copies [2], a specific signature of motor patterns. Moreover, spectrotemporal features of inner speech were decoded with significant predictive accuracy from models built from overt speech data [4].

Increasing evidence suggests that, despite the absence of vocal articulation, motor patterns form part of inner speech. However, investigation on this matter has remained challenging, due in part to the lack of behavioral output of speech imagery, which makes difficult to time-lock precise events (acoustic features, phonemes, words) to neural activity [4]. Silent reading offers a direct way to tackle this problem, since it provides us with the trajectory of the reader’s eyes along the text as a natural behavioral output [5]. Although a fair amount of behavioral experiments demonstrated that phonology affects silent reading [6,7], a quantitative study on the relation between articulatory and ocular variables is still lacking.

Here we hypothesize that articulatory simulations at single-phoneme level underlie silent reading. We tested this by a quantitative exploration of ocular and articulatory dynamics during the reading of consonant- consonant-vowel (CCV) speech structures.

Ocular trajectories basically consist of gaze fixations separated by rapid eye movements called saccades [8]. During reading, each word usually receives a few fixations, between which neighboring words may be also fixated [9]. Here we characterized reading dynamics by computing standard timing variables such as the duration of first fixations and the integration of the successive fixations on a word, which have been largely used to disentangle visual, lexical and contextual processing in reading [10].

To emphasize articulatory dynamics, focus was put on CCVs formed with plosive consonants, which require sharp movements to close the vocal tract completely. For instance, when the lips occlude the air passage, consonant *b* is produced; in this case, also the vocal folds are active. Instead, when folds are inactive, the same occlusion produce the unvoiced plosive *p* [11]. Plosives therefore involve sharp articulatory movements combining on-off folds activities [12], and sequences of plosives and consonants have also been modeled and synthesized from physical principles [13,14]. From a lexical point of view, plosive CCVs form a homogeneous subset of pseudo-words with low intra-word frequency values [15]. This combination of motor, mathematical and lexical features makes these CCVs an ideal set of stimuli to explore the relationship between silent reading and vocal articulation.

Results

Thirty native Spanish speakers read a set of consonant-consonant-vowel structures (CCV) from a computer screen. The set was formed by every combination of fricative consonants (*s, f*, *j*) and plosive consonants (*b*, *d*, *g*, *p*, *t*, *k*) ending with the vowel *a*. A few examples of these structures are *fga* (fricative-plosive), *bda* (plosive-plosive) and *jsa* (fricative-fricative), shown in Figure 1a. Intra-word frequency of each structure are displayed in Figure 1b (none is a Spanish word).

The experiment was divided into two blocks, each formed with three repetitions of the complete set of CCVs. In one block the participants read the set aloud, and in the other they were instructed to read in silence. Blocks and CCVs were randomized before presentation. For each CCV we measured the following variables, sketched in Figure 1c (videos available at Supplementary Information):

<u>Ocular</u> (aloud and silent blocks). Eye movements were recorded and the following variables were computed for each CCV: 1. the duration of the first fixation (FFD); 2. the duration of all fixations prior to passing to another CCV (First Pass Reading Time, FPRT) and 3. the total fixation time (TFT). These are standard ocular variables used to characterize cognitive processing in reading; the first two are typically considered early measures, whereas measure 3 reflects later processing stages [16,17].

<u>Phonatory</u> (aloud block). From the spectrograms of the audio records, we computed: 1. the delay Δ between the onset of visual fixation and voice onset; 2. the total duration T of the CCV, and 3. the consonantal transition τ, shown in detail in Figure 1d for the plosive-plosive combinations *bda* and *gta*. We defined τ as the interval between the release of the first occlusion, characterized by a brief voiced sound, and the release of the second one, that give rise to the vowel *a.* Finally, differences in speech rate across subjects were washed out by normalizing the transitions to the total duration of the CCV, $\tau' = \tau/T$, leaving us with the pre-phonatory variable Δ and the phonatory variable τ'.

<u>Articulatory</u> (aloud and silent blocks). Lip movement was acquired by attaching an accelerometer to the lower lips with medical tape.

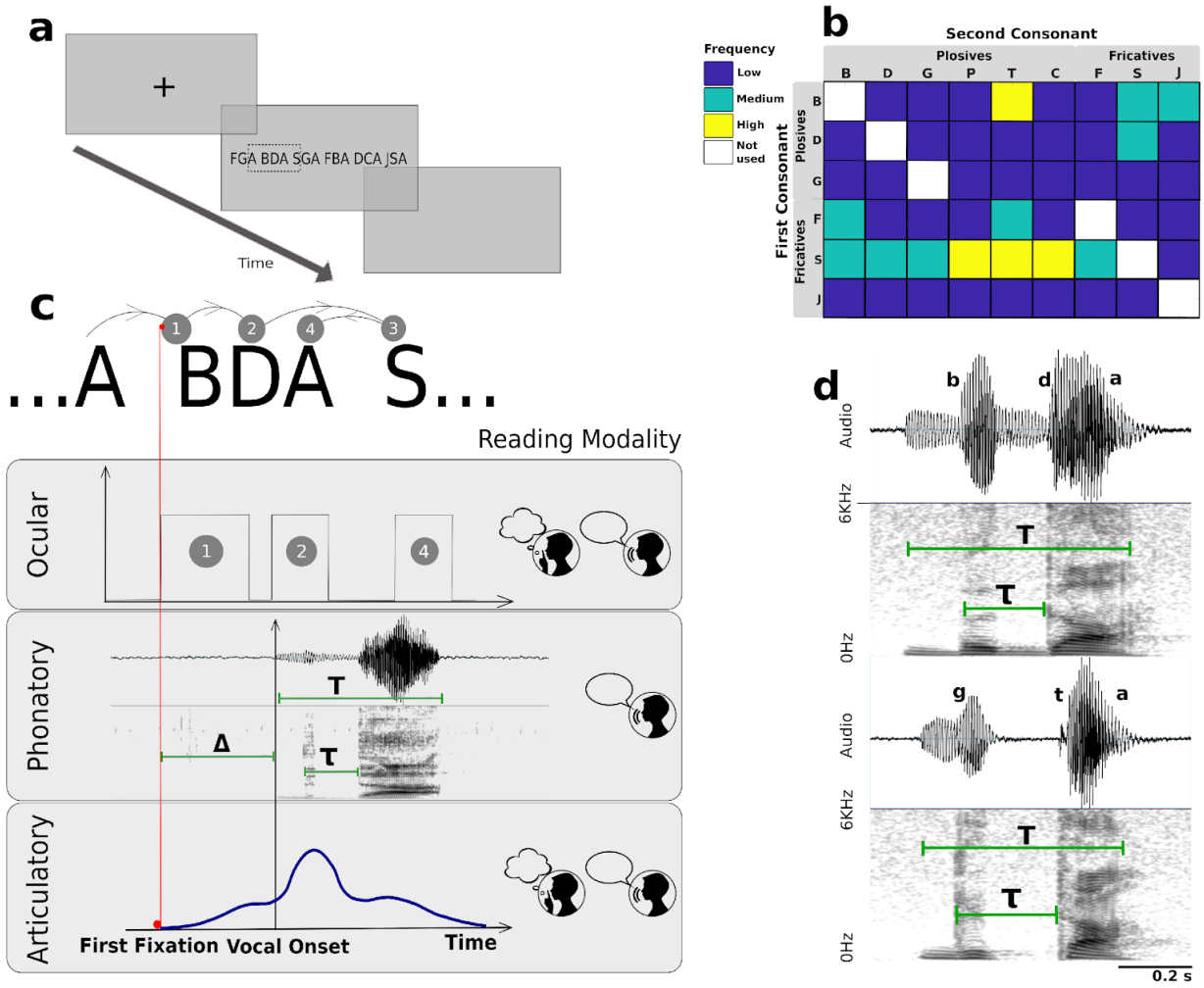


Figure 1 | Monitoring ocular, phonatory and articulatory variables during reading. a. After fixing the sight on a cross in the middle of the screen, a sequence of 6 CCVs appeared for the participant to read. The operation was repeated until three repetitions of the complete CCV set was presented. In one block, participants read aloud and in the other block they read silently. **b**. Intra-word frequency *f* of the CCVs in Spanish (in appearances per million words). The range was discretized in three levels: low (*f*<10), medium (10<*f*<50) and high (*f*>50). **c**. Three types of variables were measured during reading. Ocular: duration of first fixation (FFD=1); duration of fixations on a CCV prior to fixating a neighboring one (FPRT=1+2); and total fixation duration (TFT=1+2+4). Phonatory: We measured the delay between the first fixation and the vocal onset Δ, the transition time between consonants τ, and the total pronunciation time T (overt reading). Articulatory: Lip movement was registered by an accelerometer fixed to the lower lip. **d**. Spectrograms and phonatory variables of *bda* (voiced plosives) and *gta* (unvoiced second consonant).

Articulation is inhibited during silent reading

We first investigated the articulatory activity in both aloud and silent reading blocks. The main vocal tract articulators are the lips, the jaw and the tongue. Lip movements can be easily obtained from the exterior of the mouth, minimizing the interference of the measuring device with articulatory dynamics. Since articulators are coordinated during speech [18], we assumed that the range of lip movement is a good estimate for the range of the whole articulatory movements.

In Figure 2 we show the dynamics of the lips for the CCVs that start with the bilabial consonant *p* (*pba, pda*, *pga*, *pta* and *pca*), which requires maximal lip displacement. In Figure 2a we show the absolute value of the accelerometer signals during reading aloud, aligned to the beginning of the vocalization. We recovered the characteristic dynamics of the lips, with movement starting roughly half a second before phonation, and maximum displacement at phonation onset [19]. The fact that maxima mostly occur at vocal onset can be observed also in Figure 2b, where the signals were aligned to their maximum value instead of the vocal onset. We take advantage of this to compare the signals of the aloud block (red) with those of silent block (blue), for which we do not have the vocal onset reference. Two groups are readily identified among the silent readers (blue), one with low and the other with virtually null lips activity. Direct comparison between modalities shows that, when present, articulatory movements during silent reading are roughly two orders of magnitude smaller than those of reading aloud. These results allow us to explore how articulatory processing affects silent reading, in a context of strong inhibition of vocal articulation.

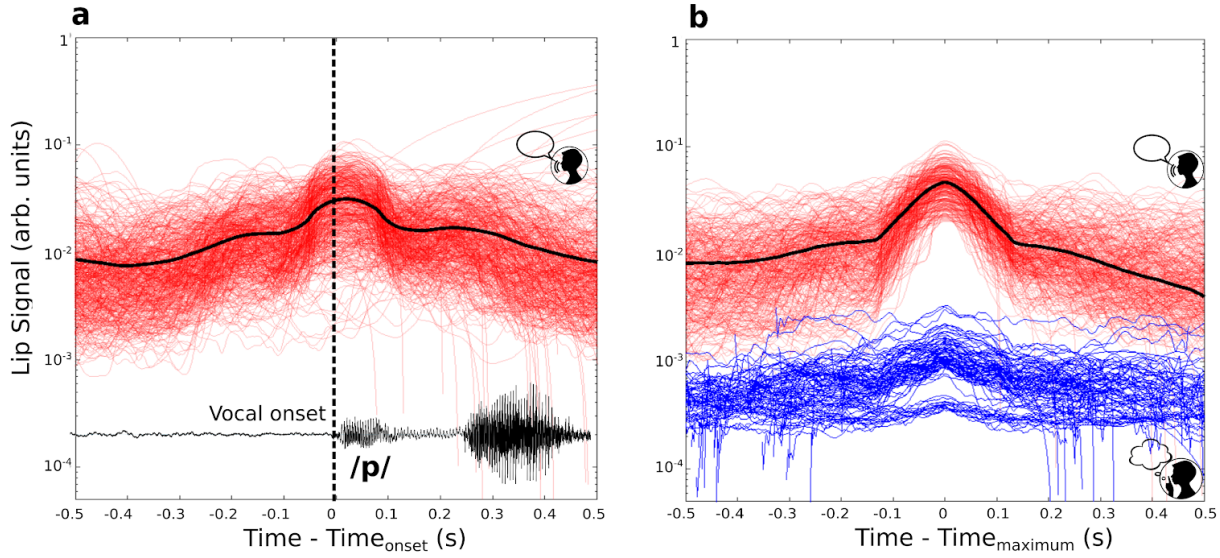


Figure 2 | Lip movement is strongly inhibited during silent reading. a. Red traces are the accelerometer signals during reading aloud, aligned at vocal onset (absolute value, smoothed by a 140 ms moving window). Single trials are represented in red and averaged in black. **b.** In order to compare ranges of articulation of the different reading modalities, the signals were realigned to the maximum. Aloud reading is displayed in red and its average in black while silent reading in blue. At maxima, traces exhibit a difference of roughly two orders of magnitude.

Silent reading is modulated by consonantal transitions

In this section we analyze the effects of phonation and memory on the ocular dynamics during reading. We focused on the CCVs formed with plosive consonants, for which articulatory movements are emphasized with respect to other consonantal families. The analyses for these other families (plosive-fricative combinations) show milder effects and can be found in Supplementary Information.

We used the consonantal transition τ' to characterize the phonation of the CCVs; since participants read each CCV six times (three times per block), the repetition number was used to test for short term memory effects. Long term memory effects were examined using the frequency of appearance of the CCVs in a Spanish corpus. Statistical tests show no effects of frequency on any visual variable (last column of Table 1)

as expected by the low frequency levels of the whole set of plosive CCVs (with the exception of BTA), as sketched in Figure 1b.

Aloud block. We analyzed the effects of phonation and memory on the ocular variables and also on the pre-phonatory variable Δ during reading aloud (Table 1, top row). For this last case, we found a significant effect of τ' on Δ, as reported in a previous work [15]. This means that the preparatory stages before phonation are longer for CCVs requiring longer consonantal transitions, which has been identified as a possible effect of articulatory processing [15]. For the ocular variables, linear regressions were conducted using consonantal transition τ' and repetition number as independent variables. A negative dependence with repetition number emerged on FPRT and TFT, and no effects were found on FFD, supporting that first fixation durations are not sensitive to habituation during reading aloud. Finally, we found no effects of transitions τ' on the ocular variables. This result was expected, given the complexity of ocular dynamics during reading aloud, in which the eyes appear to be holding in place for many fixations "so as to not get too far ahead of the voice" [20].

Aloud block vs. silent block. We next concentrated in the relationship between silent and overt reading, which is the main aim of this work. For this sake, we compared ocular data from the silent block with phonological data from the aloud block (Table 1). To compare cross-block variables, we first performed linear regressions for each ocular measure using repetition number as the independent variable. Consistent with the results obtained for the aloud block, this revealed that FPRT and TFT systematically decrease across repetitions, while FFD presents no short term memory effects. This gives us confidence that the variables that integrate the fixations on a CCV (FPRT and TFT) are affected by habituation while the duration of the first fixation FFD is not, making the latter a good candidate for articulatory processing. To test this, we compared the ocular variables of the silent block with the phonatory variable τ' of the aloud block, averaging across participants and repetitions. This is summarized in Figure 3, where we show each ocular variable as a function of consonantal transitions per CCV. A strong positive relation emerged for FFD (Figure 3a), while for TFT and FPRT the relation did not reach significance (Figure 3b and c).

Taken together, these analyses reveal that the duration of the first fixations FFD in silent reading are not affected by habituation, and are tightly correlated with the consonantal transitions τ'. This supports that ocular dynamics during silent reading are strongly modulated by phonatory features produced during the actual utterance of the speech structures.

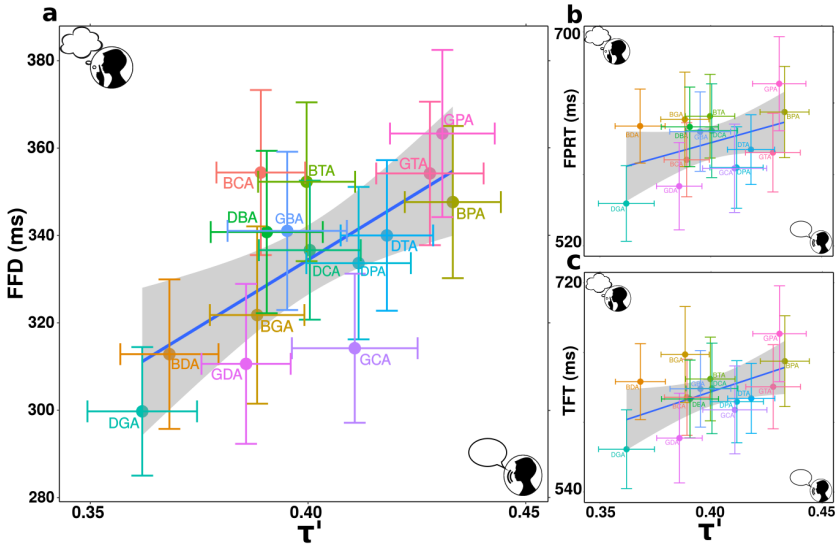


Figure 3 | Silent reading is modulated by consonantal transitions. a. First fixation durations from silent block as a function of transition τ' from aloud block. Linear regression was conducted using averaged CCVs points with its respective statistical standard error considered, showing that strong dependence between FFD and τ' ($t = 3.60, p = 3.2210^{-3}$) **b-c.** First pass reading time (FPRT) and total fixation time (TFT) as a function of τ', FPRT: ($t = 1.60, p = 0.13$); TFT: ($t = 2.30, p = 0.044$).

Consonantal transitions measure articulatory efforts

So far, our results support that silent reading is strongly modulated by the transitions τ' between consonants. Here we connect this timing variable with the motor actions needed to produce the utterance of a CCV. Like any other speech structure, producing a CCV requires two basic motor actions: the larynx needs to be abducted to control the oscillations of the vocal folds, and the vocal tract shape is reconfigured to produce the different phonemes. We explored how the consonantal transitions τ' are related to these specific motor actions.

Laryngeal effort. The vocal folds are a pair of elastic membranes located at the glottis, within the larynx. The folds can be set into oscillatory motion by the passing airflow, raising the lung pressure over a threshold that increases with glottal abduction. Within our set of CCVs, vibration of the folds is either active during the whole vocalization (*bda*, *bga*, *dba*, *dga*, *gba*, *gda*) or inactivated during the second plosive (*bca*, *bta*, *bpa*, *dca*, *dta, dpa, gca, gta, gpa*), as evidenced by the traces of the fundamental frequency in the spectrograms of Figure 1d. To produce the CCVs in the second group, the glottis needs to be actively abducted to prevent vibration during the middle unvoiced plosive, which involves a devoicing effort. Consonantal transitions reflect this, with a smaller mean duration for the former group, $\tau' = (38.2 \pm 0.3)\,10^{-2}$ than for the latter, $\tau' = (41.3 \pm 0.3)\,10^{-2}$ (t-test: $t = -4.99, p < 10^{-4}$).

Tract effort. A CCV formed by plosives is produced by two successive occlusions that occur while the tract evolves towards the vowel *a*. This produces slight differences in the articulatory effort when consonants are interchanged. For instance, *bda* requires less effort than *dba* because in the latter, the lips' closure for the *b* occurs closer to the vowel *a*, for which the mouth is fully opened, requiring a larger vocal tract deformation. These slight vocal tract asymmetries are also reflected by consonantal transitions, with $\tau' = (37.3 \pm 0.7)\,10^{-2}$ for the group (*bda*, *bga*, *dga*) and $\tau' = (39.0 \pm 0.7)\,10^{-2}$ for the group with interchanged consonants (*dba*, *gba, gda*), with a trend towards significance ($t = 1.79, p = 0.074$). Beyond this subtlety, the main variations in the tract effort arise from the anatomical differences between plosives: for instance, producing a *g* requires displacing the body of the tongue, while for a *d* only the tip is shifted. These anatomical differences can be accounted for using mathematical functions $A(x,t)$ for the cross section of the vocal tract along its length x from the entrance to the mouth.

In a previous work [15] we capitalized on this mathematical description, joining together the laryngeal and vocal tract components in a single equation for the vocal effort E that reads:

$$E = E_0 + \int_0^T \int_0^L \left| A(x,t) - A_\Omega(x) \right| dx dt \qquad (1)$$

The first term refers to the laryngeal component that takes a value $E_0 > 0$ when a phoneme is devoiced, and $E_0 = 0$ otherwise. The second term accounts for the elastic forces produced by deformations of the

vocal tract, from a neutral shape $A_\Omega(x)$ to a general shape $A(x)$, as sketched in Figure 4a. The tract effort is obtained by integrating the section $A(x,t)$ along the duration T of a CCV, from the vocal tract entrance at $x = 0$ to mouth exit at $x = L$. In this way, equation 1 provides an estimate of the total laryngeal and vocal tract articulatory actions needed to produce our speech structures.

To compute the efforts of each CCV, we used vocal tract functions $A(x,t)$ and devoicing efforts E_0 from anatomical data as reported elsewhere [15,21,22]. In Figure 4b we show the estimated efforts E for each CCV as a function of τ' (averaged over participants and repetitions). This result supports that consonantal transitions are a good estimate for the vocal effort needed to produce a CCV at laryngeal and vocal tract levels.

Taking together the analyses of the previous sections, our results show that silent reading involves the processing of precise articulatory features. First fixations during silent reading are lengthened when the CCVs require a greater laryngeal and/or articulatory effort to be pronounced, even when this implies devoicing a single consonant or changing the point of occlusion in the vocal tract.

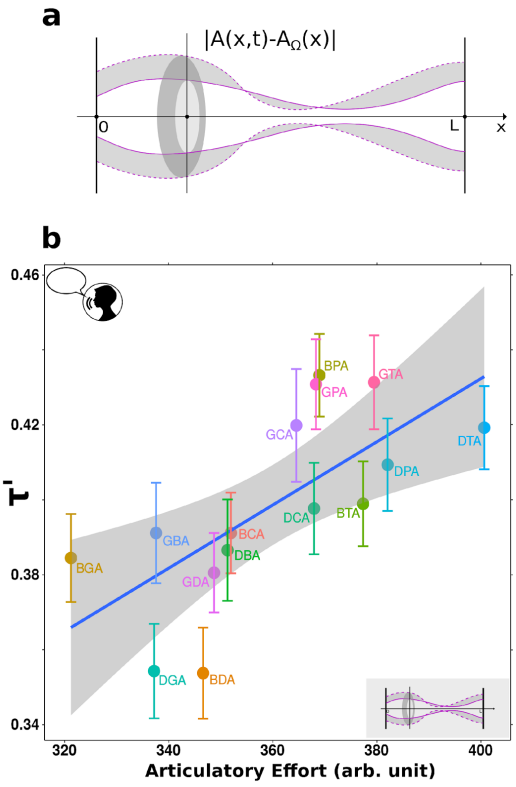


Figure 4 | Computing total vocal effort from consonantal transitions. a. The shape of the vocal tract is described by its cross section $A(x,t)$, where x is the distance from the vocal tract entrance ($x = 0$) to the exit of the mouth ($x = L$). The articulatory effort needed to pronounce a CCV can be estimated from the deformations of the vocal tract (shadowed region) with respect to a relaxed configuration (dotted line). **b.** The transitions τ' measured on the recorded CCVs averaged across participants, exhibit a positive relation with the modeled articulatory effort computed with Equation 1 ($t = 3.4, p = 4.7\ 10^{-3}$).

Conclusions and Discussion

The faculty of language involves the ability to switch between inner and actual speech, offering a unique opportunity to study the differences in motor processing between these operations. We capitalized on this faculty to explore the articulatory signatures during aloud and silent reading. Our main findings are that 1. the duration of the first fixations on the CCVs in silent reading are strongly correlated with the duration of the consonantal transitions when the CCVs are read aloud, and 2. consonantal transitions measure the vocal effort needed to pronounce the CCVs. Together, these results support that short pseudowords are processed at single-phoneme articulatory level during silent reading.

Two theories have accounted very differently on the role of the motor system during inner speech tasks. The abstraction theory supports that inner speech only activates abstract linguistic representations, independently from any articulatory mental simulation [23,24]. On the other hand, the motor simulation theory [25,26] describes inner speech as an activity that involves a similar motor processing than overt speech, including articulatory detail. Halfway between both theories, Oppenheim and Dell proposed a flexible abstraction hypothesis, in which inner speech operates at two levels: an abstract processing level, and one that incorporates a lower-level articulatory planning [27].

Recent investigations accumulate evidence towards motor simulations; Whitford et al [2] showed that speech imagery is associated with an efference copy with detailed auditory properties, treating it as a kind of action. In the same direction, a groundbreaking result was reported by Anumanchipalli and colleagues [28], who mounted a neural decoder based on articulatory dynamics capable of synthesizing speech when a participant silently mimed sentences. Investigation on inner speech has direct implications in the development of devices to restore speech [29], and also in other fields as psychiatry, where failure in the monitoring of inner speech has been associated with verbal hallucinations in schizophrenia [30].

However, not so much attention has received the motor processing during silent reading. Previous research showed that silent reading activates phonological representations in the brain [6], and also that global aspects of overt speech affects the dynamics of silent reading [7]. In line with this research, our work is a quantitative in-depth study of the articulatory signatures during silent reading. We have shown that slight motor variations are sufficient to affect the reading dynamics and indicate that this task, as speech imagery, is assisted by precise vocal simulations.

Materials and Methods

Participants

Thirty native Spanish speakers (13 females and 17 males, age range 22–33, mean age 27), undergraduate and graduate students at the University of Buenos Aires with normal vision, no speech impairments and fluent reading skills completed the experiment. Participants signed a written consent form. All the experiments described in this paper were approved by the CEPI Ethics Committee of the Hospital Italiano de Buenos Aires, qualified by ICH (FDA-USA, European Community, Japan) IRb00003580, and all methods were performed in accordance with the relevant guidelines and regulations.

Stimuli and tasks

The experiment consisted of two blocks separated by a three minute break, each involving a single reading modality. Before the aloud reading block, participants were instructed to say the pseudo-words as they appeared on the screen. Before the silent reading block, participants were instructed to read the sequences silently, as they normally do when reading a book. The order of the blocks was randomized.

The experimental design was identical for each block: a screen containing a sequence of 6 randomized consonant-consonant-vowel (CCV) structures was presented right after the participant fixated a red point (Figure 1a). Once the CCVs were read, the participant pressed a key and the process restarted until completing the set (videos are available as Supplementary Materials).

The CCVs were built using fricative and plosive consonants followed by the vowel *a*. We used the most common Spanish fricatives *f*, *s* and *j* (pronounced as **f**ace, **s**tand, and the scottish lo**ch**) and all the Spanish voiced plosives *b*, *d, g* (**b**ay, **d**ye and **g**ray) and unvoiced plosives *p, t, c* (**p**ay, **t**ie and **c**ray). Examples of these structures are *fsa* (fricative-fricative), *sda* (fricative-plosive) and *bta* (plosive-plosive). We excluded CCVs that repeat the same consonant, which makes a total number of combinations of $9 \times 8 = 72$ CCVs. Here we analyzed the 15 CCVs formed by a voiced plosive (*b*, *p*, *d*) followed by any other plosive of the set (*t*, *g, c, b*, *p*, *d*). The reason is that in these combinations, vocal folds vibrate the initial voiced plosives, allowing us to determine precisely the beginning of the CCV in the spectrogram. Unvoiced plosives, on the contrary, leave no sound or spectrogram traces and were excluded from the analysis. Voiced plosive-plosive combinations were repeated three times per block (15 CCVs 3 repetitions = 45), while all other 57 CCVs were repeated randomly two or three times -adding 135 repetitions- to make a total of 180 stimuli (6 CCVs per screen 30 trials) per block. Data was recorded for 10800 CCV samples (30 subjects 2 blocks 180 CCVs). In order to avoid screen edges effects on the ocular variables, the first and last CCVs on the screen were discarded (Figure 1a). The final database was composed by 7200 CCV samples.

The frequency *f* of each combination of consonants was computed from a large Spanish corpus [31], as the intra-word appearances per million words. Since the range of frequency values was large ($0 \le f \le 10^5$), the range was discretized in three levels: low ($f \le 10$), medium ($10 \le f \le 50$) and high ($f \ge 50$). None of the CCVs are Spanish words. Courier New monospaced font was used to ensure a fixed-width of 24 pixels for every character. Stimuli were presented in black over a white background.

Eye tracking data

Visual trajectories were acquired through a desktop-mounted, video-based eye tracker (EyeLink II; SR Research Ltd., Kanata, Ontario, Canada) at a sampling frequency of 1 kHz in binocular mode (nominal average accuracy 0.5°, space resolution 0.01° RMS). The stimuli were presented on a 19 in. monitor model Samsung SyncMaster 997 MB, at an eye–monitor distance of 50 cm. Eye position was recorded at a resolution of 1024 × 768 px in all tasks. The head was stabilized with a forehead rest.

Eye movements during reading include saccades, which are rapid eye movements, and fixations, where the eyes remain relatively still for about 200–300 ms. In this experiment we computed three of the most frequent variables used for measuring processing time [20]: 1. First Fixation Duration (FFD), 2. First Pass Reading Time (FPRT), which is the sum of all fixation durations on the word before any other word is fixated, and 3. Total Fixation Time (TFT), the sum of all fixations durations on a word.

Acoustic and articulatory data

Sound was recorded with a commercial microphone placed at 0.3 m from the head of the participant. Lip movement was recorded using an accelerometer ADXL335 fixed to the lower lip with medical paper tape.

Recordings were made in the direction of maximum amplitude of lip movement, along the perpendicular axis of the accelerometer. Sound and articulatory signals were acquired with a DAQ (USB-1608FS-Plus, Measurement Computing, Norton, Massachusetts), which ensured synchronization of both inputs and a maximum onset delay of 50 ms. Psychtoolbox library from MATLAB was used to synchronize the eye tracker and DAQ system to the computer clock.

For each CCV, we measured the following three timing variables from the audio records:

a. The delay Δ between the first fixation and the vocalization onset. Vocal onsets were established at the appearance of the spectral signature of the initial consonant (noisy spot for fricatives and fundamental frequency for voiced plosives).
b. The duration T of the uttered CCV.
c. The transition τ during the reconfiguration of the vocal tract from the first to the second consonant. We followed the procedure described in [15] to define the transitions τ in the time- frequency domain. For voiced plosive-plosive combinations, as shown in Figure 1d, τ goes from the release of the first plosive, characterized by a voiced sound structure, to the release of the second one into the vowel *a;* for plosive-fricative combinations, τ is the interval bounded by the brief broad-band plosive noise and the purely noisy fricative spot, as shown for *gfa* in Supplementary Figure S1a; for fricative-plosive combinations, the transition is the interval between the abrupt end of the fricative and the release of the plosive into the vowel, as shown for *fca* in Supplementary Figure S2a.

To wash out the speech rate differences between speakers, normalized transitions $\tau' = \tau/T$ were used throughout this work.

From the recorded 2700 voiced plosive-plosive combinations, we discarded those that were not fixated with the eyes, the mispronounced ones and those that contained more than one attempt to produce a CCV. We have also excluded 2 participants who mispronounced more than 45% of the CCVs throughout the experiment, leaving a dataset of 2415 CCVs. We then discarded repetition which presented fixation durations (FFD, FPRT and TFT), delay times Δ or transitions τ' out of the 95% confidence interval ($\underline{x} \pm 2s$) from the participant's mean value of each variable. A final dataset of ocular, acoustic and articulatory data of 2178 CCVs from 28 participants was used to perform the analyses throughout this work.

Vocal tract model

During a sequence of vowels and *n* plosive consonants, the vocal tract area $A(x,t)$ can be mathematically described by [21]:

$$A(x,t) = \tfrac{\pi}{4}\left[\Omega(x) + q_1(t)\,\phi_1(x) + q_2(t)\,\phi_2(x)\right]^2 \prod_{k=1}^{n}\left[1 - c_{k_n}(x)\,m\left(t - t_{k_n}\right)\right] \qquad (2)$$

The squared factor represents the vowel substrate, with empirical functions Ω, ϕ_1 and ϕ_2 obtained from an orthogonal decomposition calculated from MRI anatomical data. Any sequence of vowels is generated by the evolution of q_1 and q_2. Plosive consonants are generated through the bell-shaped functions c_k and m. These functions reach the value 1 around $x = x_k$ and $t = 0$ respectively, producing a specific occlusion $A(x_k, t_k) = 0$. The specific features of each plosive is represented by the spatial functions c_k that characterize the anatomy of the occlusion and also by the time function m that represents the temporal dynamics of the occlusion (activation-deactivation).

Acknowledgements

We thank Juan Kamienkowski for his time and willingness during the experiments. The research reported in this work was partially funded by the National Scientific and Technical Research Council (CONICET), the University of Buenos Aires (UBA).

References

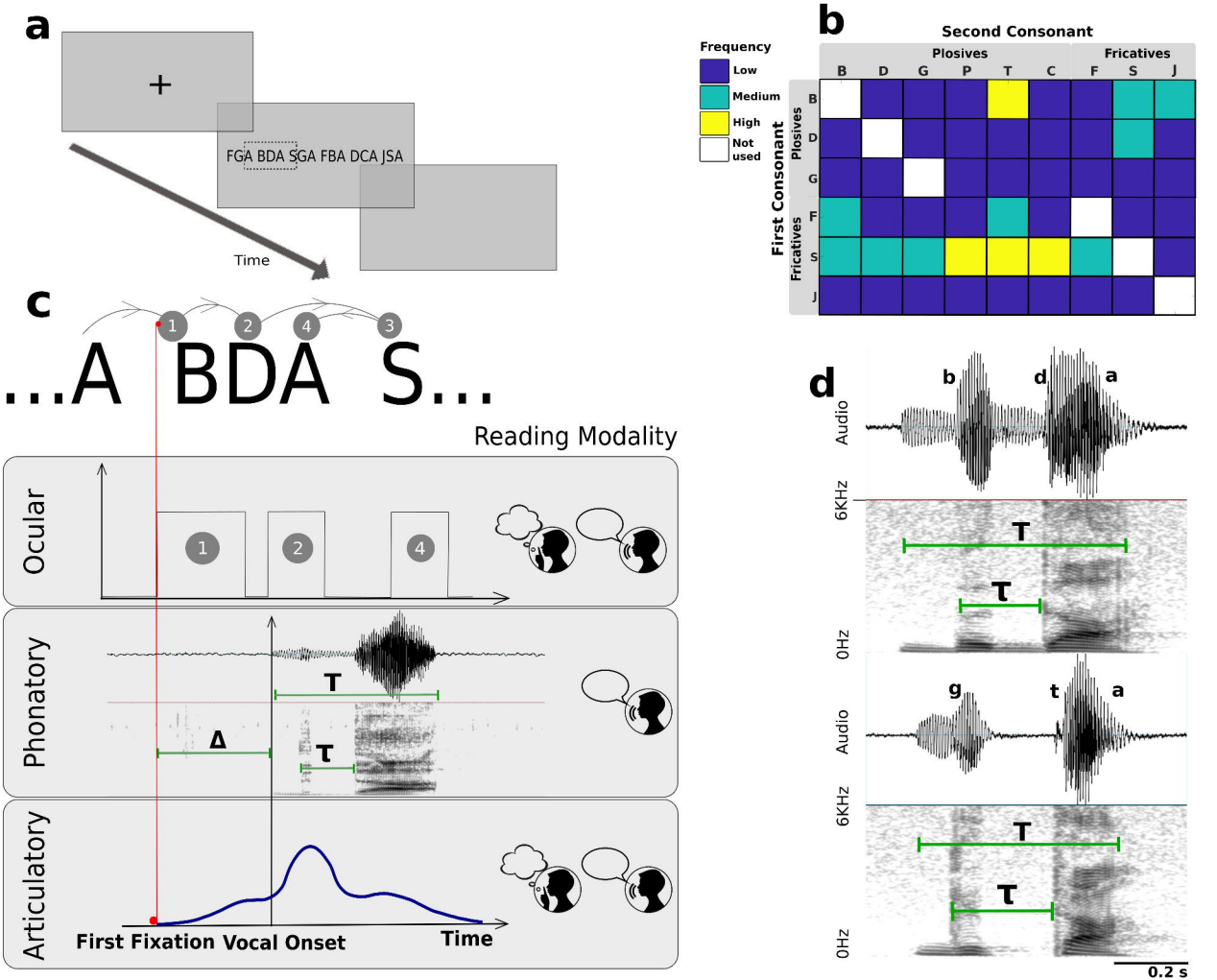
1. Mcguigan AFJ, Camacho EO, Hardyck CD, Petrinovich LF, Delbert W. Feedback of Speech Muscle Activity during Silent Reading: Rapid Extinction. Science (80-). 1967;157: 579–581.
2. Whitford TJ, Jack BN, Pearson D, Griffiths O, Luque D, Harris AWF, et al. Neurophysiological evidence of efference copies to inner speech. Elife. 2017; 1–23. doi:10.7554/eLife.28197.001
3. Cheung C, Hamilton LS, Johnson K, Chang EF. The auditory representation of speech sounds in human motor cortex. Elife. 2016;5: 1–19. doi:10.7554/eLife.12577
4. Martin S, Iturrate I, Millán J del R, Knight RT, Pasley BN. Decoding inner speech using electrocorticography: Progress and challenges toward a speech prosthesis. Front Neurosci. 2018;12: 1–10. doi:10.3389/fnins.2018.00422
5. Laubrock J, Kliegl R. The eye-voice span during reading aloud. Front Psychol. 2013;6: 1–19. doi:10.3389/fpsyg.2015.01432
6. Abramson M, Goldinger SD. What the reader's eye tells the mind's ear: Silent reading activates inner speech. Percept Psychophys. 1997;59: 1059–1068. doi:10.3758/BF03205520
7. Filik R, Barber E. Inner speech during silent reading reflects the reader's regional accent. PLoS One. 2011;6. doi:10.1371/journal.pone.0025782
8. Dehaene S. Reading in the brain: The new science of how we read. Penguin; 2009.
9. Fernández G, Shalom DE, Kliegl R, Sigman M. Eye movements during reading proverbs and regular sentences: The incoming word predictability effect. Lang Cogn Neurosci. 2014;29: 260–273. doi:10.1080/01690965.2012.760745
10. Kliegl R, Grabner E, Rolfs M, Engbert R. Length, frequency, and predictability effects of words on eye movements in reading. Eur J Cogn Psychol. 2004;16: 262–284. doi:10.1080/09541440340000213
11. Story BH, Titze IR, Hoffman EA. Vocal tract area functions from magnetic resonance imaging. J Acoust Soc Am. 1996;100: 537–54. Available: http://www.ncbi.nlm.nih.gov/pubmed/8675847
12. Story BH, Bunton K. Relation of vocal tract shape, formant transitions, and stop consonant identification. J Speech Lang Hear Res. 2010;53: 1514–28. doi:10.1044/1092-4388(2010/09-0127)
13. Assaneo MF, Nichols JI, Trevisan MA. The Anatomy of Onomatopoeia. Sigman M, editor. PLoS One. 2011;6: e28317. doi:10.1371/journal.pone.0028317
14. Assaneo MF, Sitt J, Varoquaux G, Sigman M, Cohen L, Trevisan MA. Exploring the anatomical encoding of voice with a mathematical model of the vocal system. Neuroimage. Elsevier B.V.; 2016;141: 31–39. doi:10.1016/j.neuroimage.2016.07.033
15. Taitz A, Shalom DE, Trevisan MA. Vocal effort modulates the motor planning of short speech structures. Phys Rev E. American Physical Society; 2018;052406: 1–7. doi:10.1103/PhysRevE.97.052406
16. Clifton Jr C, Staub A, Rayner K. Eye movements in reading words and sentences. Eye Movements. Elsevier; 2007. pp. 341–371.
17. Vasishth S, von der Malsburg T, Engelmann F. What eye movements can tell us about sentence comprehension. Wiley Interdiscip Rev Cogn Sci. Wiley Online Library; 2013;4: 125–134.
18. Goldstein L, Byrd D, Saltzman E. The role of vocal tract gestural action units in understanding the evolution of phonology. Action to Language via the Mirror Neuron System. Cambridge University Press; 2006. pp. 215–249.
19. Bouchard KE, Mesgarani N, Johnson K, Chang EF. Functional organization of human sensorimotor cortex for speech articulation. Nature. Nature Publishing Group; 2013;495: 327–32. doi:10.1038/nature11911

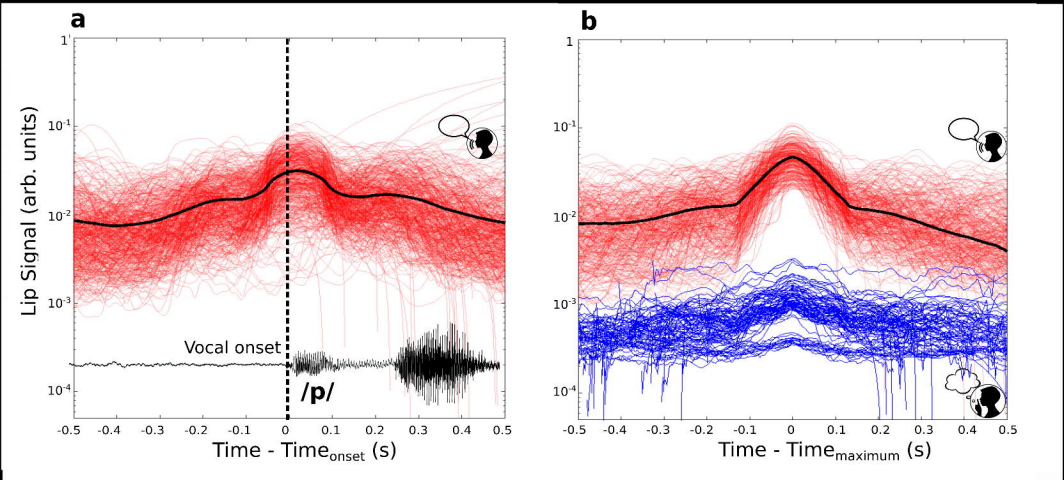
20. Rayner K. Eye Movements in Reading and Information Processing : 20 Years of Research. Psychol Bull. 1998;124: 372–422.
21. Story BH. A parametric model of the vocal tract area function for vowel and consonant simulation. J Acoust Soc Am. 2005;117: 3231. doi:10.1121/1.1869752
22. Story BH. Phrase-level speech simulation with an airway modulation model of speech production. Comput Speech Lang. Elsevier Ltd; 2013;27: 989–1010. doi:10.1016/j.csl.2012.10.005
23. MacKay DG. Constraints on theories of inner speech. Auditory imagery. Psychology Press; 2014. pp. 133–162.
24. Dell GS, Repka RJ. Errors in Inner Speech. Experimental Slips and Human Error. Springer; 1992. pp. 237–262. doi:10.1007/978-1-4899-1164-3_10
25. Levelt WJM. Speaking: From intention to articulation. MIT press; 1993.
26. Postma A, Noordanus C. Production and detection of speech errors in silent, mouthed, noise-masked, and normal auditory feedback speech. Lang Speech. SAGE Publications Sage UK: London, England; 1996;39: 375–392.
27. Oppenheim GM, Dell GS. Motor movement matters: the flexible abstractness of inner speech. Mem Cogn. 2010;38: 1147–1160. doi:10.3758/MC.38.8.1147.Motor
28. Anumanchipalli GK, Chartier J, Chang EF. Speech synthesis from neural decoding of spoken sentences. Nature. Springer US; 2019;568: 493–498. doi:10.1038/s41586-019-1119-1
29. Assaneo MF, Butavand DR, Trevisan MA, Mindlin GB. Discrete anatomical coordinates for speech production and synthesis. Front Commun. 2019;4: 1–13. doi:10.3389/fcomm.2019.00013
30. Brébion G, Stephan-Otto C, Ochoa S, Roca M, Nieto L, Usall J. Impaired self-monitoring of inner speech in schizophrenia patients with verbal hallucinations and in non-clinical individuals prone to hallucinations. Front Psychol. 2016;7: 1–12. doi:10.3389/fpsyg.2016.01381
31. Cuetos F, Glez-Nosti M, Barbón A, Brysbaert M. SUBTLEX-ESP: Spanish word frequencies based on film subtitles. Psicológica. 2011;32: 133–143. doi:10.4090/juee.2008.v2n2.033040

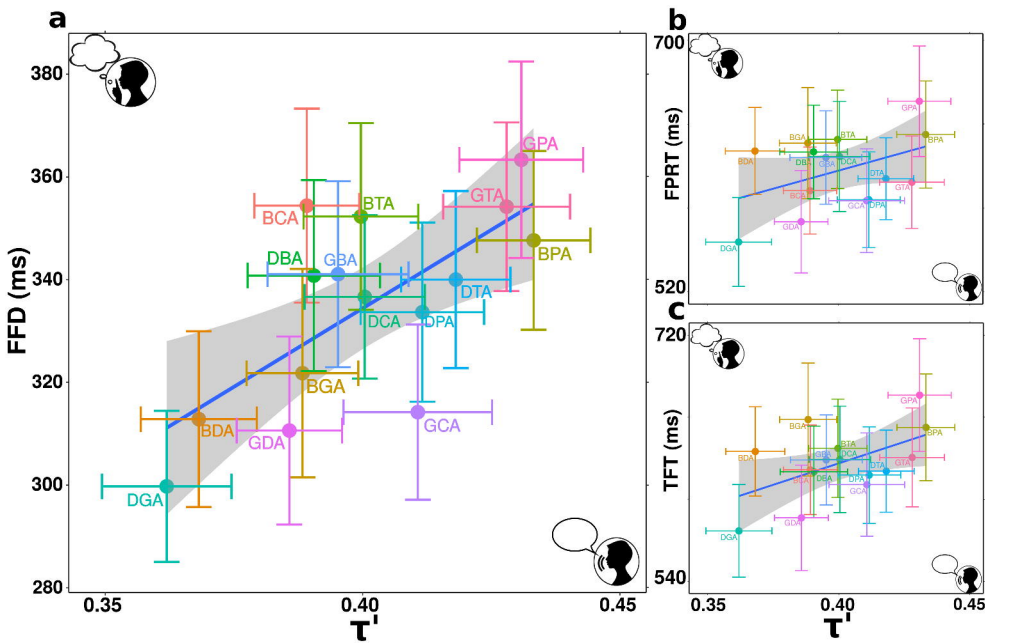
Tables

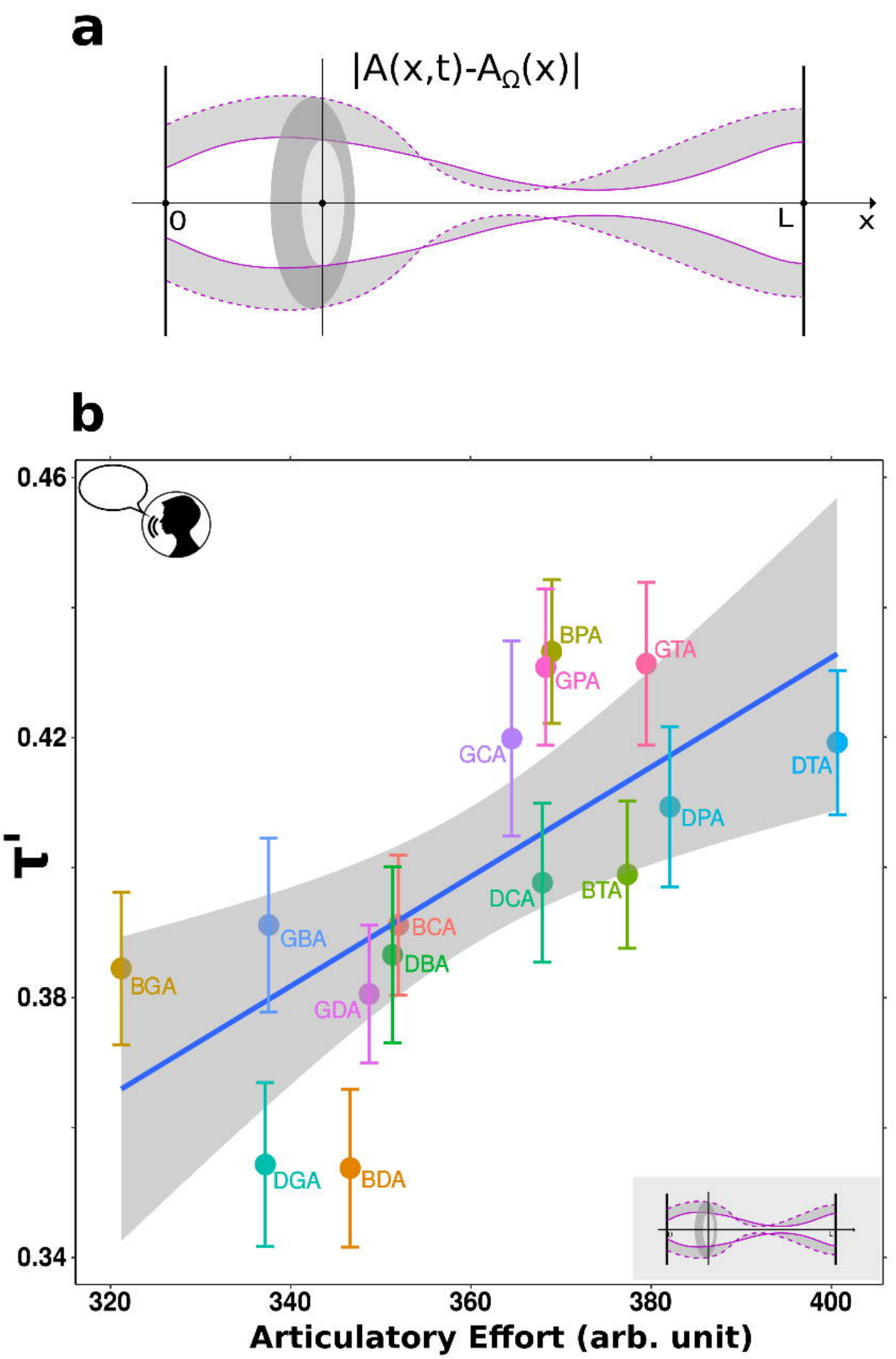
		Consonantal transition τ'		Repetition Number		Intra-word frequency	
		t	**p**	**t**	***p***	**F**	***p***
	Vocal onset delay Δ	**5.54**	**< 10^{-4}**	-1.74	0.082	1.32	0.25
	Total fixation time TFT	2.13	0.034	**-5.60**	**< 10^{-4}**	1.82	0.18
	First pass FPRT	1.78	0.076	**-5.16**	**< 10^{-4}**	1.25	0.26
	First fixation FFD	-0.094	0.93	0.59	0.56	0.33	0.57
	Total fixation time TFT	2.23	0.044	**-9.15**	**< 10^{-4}**	0.10	0.75
	First pass FPRT	1.60	0.13	**-9.01**	**< 10^{-4}**	0.39	0.53
	First fixation FFD	**3.6**	**3.22 10^{-4}**	0.84	0.40	1.25	0.26

Table 1 | Effects of phonology and memory on ocular variables during reading. An ANOVA test revealed no significant effect of intra-word frequency on any of the variables. **Aloud block**. A multiple linear regression was conducted for the onset delay Δ and for each of the ocular measures, with consonantal transition and repetition number as independent variables. **Silent Block.** To compare variables across blocks, two different tests were performed. Short term memory effects on the ocular measures was accounted for by performing a linear regression test with repetition number as the independent variable. Ocular variables (silent block) and consonantal transitions (aloud block) were averaged across participants and repetitions before performing a linear regression test weighted by the statistical errors (gray background).


a
+
FGA BDA SGA FBA DCA JSA
Time
b
Second Consonant
Plosives
Fricatives
B D G P T C F S J
Frequency
Low
Medium
High
Not used
First Consonant
Plosives
Fricatives
B D G T S J
c
1 2 4 3
...A BDA S...
Reading Modality
Ocular
1 2 4
Phonatory
T
Δ
τ
Articulatory
First Fixation Vocal Onset
Time
d
b d a
Audio
6KHz
0Hz
T
τ
g t a
Audio
6KHz
0Hz
T
τ
0.2 s


a
b
Lip Signal (arb. units)
1
10^{-1}
10^{-2}
10^{-3}
10^{-4}
Vocal onset
/p/
-0.5 -0.4 -0.3 -0.2 -0.1 0 0.1 0.2 0.3 0.4 0.5
Time - $Time_{onset}$ (s)
Time - $Time_{maximum}$ (s)


a
FFD (ms)
380
360
340
320
300
280
0.35
0.40
0.45
τ'
GPA
BCA
BTA
GTA
BPA
DBA
GBA
DTA
DCA
DPA
BGA
BDA
GDA
GCA
DGA
b
FPRT (ms)
700
520
c
TFT (ms)
720
540


a
|A(x,t)-A_Ω(x)|
O
L
x
b
0.46
0.42
0.38
0.34
Γ
BPA
GPA
GTA
GCA
DTA
DPA
DCA
BTA
BCA
GBA
DBA
BGA
GDA
DGA
BDA
320
340
360
380
400
Articulatory Effort (arb. unit)