Journal of Memory and Language 121 (2021) 104283

Contents lists available at ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

Language modality and temporal structure impact processing: Sign and speech have different windows of integration

Chiara Luna Rivolta [a,b,*], Brendan Costello [a], Manuel Carreiras [a,b,c]

[a] Basque Center on Cognition, Brain and Language (BCBL), Spain
[b] University of the Basque Country (UPV-EHU), Spain
[c] Ikerbasque, Basque Foundation for Science, Spain

ARTICLE INFO

Keywords:
Sign language
Locally time-reversed speech
Temporal processing
Intelligibility
Language modality

ABSTRACT

Language comprehension depends on the ability to temporally process the periodic structure of the language signal. In this study we investigate temporal processing of Spanish Sign Language (LSE), isolating the specific contribution of signal modality through a comparison with spoken Spanish and the contribution of linguistic status by comparing language with a non-linguistic temporally structured visual signal. Twenty-three highly proficient hearing users of LSE performed an intelligibility task with these three types of materials, manipulated with different levels of temporal distortion. The results show that the distortion differently affects the intelligibility of these signals. Spanish is characterized by a threshold of temporal distortion, beyond which intelligibility rapidly decreases and is almost completely lost. Conversely, in LSE and the visual non-linguistic task greater temporal distortion led to a gradual and constant reduction in intelligibility with no clear threshold. LSE is more resilient to temporal manipulation compared to the visual non-linguistic signal: participants' performance never drops below 50% even with the most severe distortion. Overall, these findings suggest that the temporal processing of language arises from the complex interaction between the properties of the sensory system and the special characteristics of the language signal.

Introduction

Temporal structure of language

In the field of psycholinguistics a great effort has been devoted to characterizing the temporal properties of speech and converging results have highlighted that language is a quasi-periodic signal. This temporal regularity is evident in both behavioural and neurophysiological measures, and across different domains. Measurement of speech articulator movements has revealed temporal regularity (Walsh & Smith, 2002) and, not incidentally, the acoustic speech envelope shows a peak in the same frequencies, which fall within the delta (<4 Hz) and theta (4–8 Hz) frequency bands, as they are known in the electrophysiological literature (Ding et al., 2017). These regularities are reflected in the perceptual domain as well: during speech perception neural oscillations entrain mainly to those frequency bands present in the acoustic envelope (Meyer, 2018). Interestingly, this temporal periodicity is relatively stable across different speakers and spoken languages, pointing towards the existence of a common neural and cognitive mechanism devoted to language processing.

The temporal regularity which characterizes the speech signal seems to play an extremely important role in language perception and comprehension: a quasi-periodic structure supports the efficient parsing and decoding of linguistic information from the acoustic signal. A failure in perceiving and processing this periodicity has been linked with reading deficits and poor language comprehension (Doelling, Arnal, Ghitza, & Poeppel, 2014). Alternatively, these temporal regularities found in language can be ascribed to the properties of the acoustic sensory channel (Cummins, 2012; Samuel, 1991). According to this view, language periodicity does not reflect any specifically linguistic segmentation of the signal: listeners pick up temporally structured patterns in any type of acoustic signal in order to process it (Samuel, 2020). A way to examine this possibility is to test the perception of language but also of non-linguistic signals that show a temporally organized structure.

Language is not limited to the oral channel. When the acoustic modality is impaired, for example in the deaf and hard-of-hearing population, the visual modality is recruited. Signed languages are fully-developed, natural languages with a complex syntactic and semantic

* Corresponding author at: Basque Center on Cognition Brain and Language (BCBL), Paseo Mikeletegi, 69, 20009 San Sebstián, Spain.
E-mail address: crivolta@bcbl.eu (C.L. Rivolta).

https://doi.org/10.1016/j.jml.2021.104283
Received 10 December 2020; Received in revised form 11 August 2021; Accepted 13 August 2021
Available online 27 August 2021
0749-596X/

structure (MacSweeney, Capek, Campbell, & Woll, 2008). Linguistic information is delivered in the visual modality through the use of different articulators, both manual and non-manual (including the torso, head, eye-gaze, eyebrows and mouth). The spatial domain has a predominant role in signed languages, and the temporal structure serves these spatial characteristics. In spoken languages information is presented, and therefore processed, in a sequential fashion due to the mechanical constraint imposed by the speech articulators and the acoustic channel. Conversely, in signed languages the articulators have a certain degree of independence in motion, allowing the parallel presentation of multiple pieces of linguistic information (Sandler, 2018). Within the purely manual component of the signal, phonological theories of sign language typically identify three main phonological parameters composing the sign: the handshape, the location of the sign (on the signer's body or in the signing space) and the movement performed by the hand in the space (Brentari, 1998; Sandler, 2011; Stokoe & Marschark, 2005). These parameters are realized simultaneously during signing and the way they spatially develop across time is quite different: while movement constantly changes during signing, handshape and location are more stable and less susceptible to change within a given sign. Although there is a large corpus of studies investigating temporal structure and processing of spoken languages (Poeppel & Assaneo, 2020), this topic is still understudied in signed languages.

Two important factors differentiate the temporal structure of signed and spoken languages. Firstly, signed languages are characterized by the use of bigger articulators compared to those of spoken language; articulator movements therefore take more time to be completed. Several studies have measured sign duration. Despite some discrepancy in the results due to a lack of agreement on the precise definition of sign boundaries, the findings converge to show that sign duration is about twice the duration of a monosyllabic word (Grosjean, 1977; Klima, E. S., & Bellugi, 1979; Wilbur, 2009). At the same time, sentence duration seems to be very similar across spoken and signed languages (Bellugi & Fischer, 1972), driven by the fact that in sign language fewer signs are needed to convey the same linguistic content due to a parallel presentation of the information. Another important difference between the temporal structure of spoken and signed languages comes from the nature of the sensory modalities employed. While the acoustic modality is characterized by a high temporal resolution, the visual modality favours spatial over temporal information (Meier, 2002). Overall, both physical and linguistic information is organized on a larger time scale in signed languages compared to spoken languages.

Locally time-reversed speech paradigm

One method to investigate the role of temporal structure in language comprehension is to examine to what degree temporal distortion of the speech signal can impair intelligibility. The locally time-reversed speech paradigm has been employed in several experiments with different spoken languages (Greenberg & Arai, 2001; Steffen, A., & Werani, 1994; Stilp, Kiefte, Alexander, & Kluender, 2010). In this paradigm the linguistic signal is first divided into windows of fixed duration and then each window is reversed while the global order of the windows is kept. Fig. 1 shows an example of this manipulation applied to Spanish Sign Language (LSE). Participants hear (or view) sentences distorted in this manner and repeat what they have understood. How intelligible the distorted signal is depends on the size of the reversal window: larger windows create higher levels of distortion and are associated with lower accuracy in the repetition task. A *meta*-analysis of several studies employing this paradigm (Ueda, Nakajima, Ellermeier, & Kattner, 2017) showed that the intelligibility pattern is consistent across different spoken languages, even when they are characterized by different timing patterns (e.g. syllable-, mora- or stressed-based). Intelligibility starts to decrease when the reversal window is approximately 40 ms, drops under 50% when the window is between 60 and 70 ms and for windows of 100 ms or longer speech comprehension is almost completely lost.

To our knowledge only one study has adapted this paradigm to a sign language. Hwang (Hwang, 2011) investigated the locally time-reversed paradigm with deaf signers (native and late-learners) of American Sign Language (ASL). The results confirmed the general tendency of language intelligibility to decrease with longer reversal windows, but also revealed important differences from the pattern associated with spoken languages. Firstly, the temporal scale of the signed signal is slower than that of speech, and this was factored into the experimental design by using larger reversal windows for sign language. Intelligibility of the ASL stimuli was characterized by a slow, constant decrease as the size of the reversal window increased. The decrease stops at reversal windows of approximately 500 ms, at which point intelligibility starts to level off at 50% accuracy. A comparison between native and late learners of ASL showed an effect of age of acquisition: the intelligibility level of late-learners was lower than that of native signers across all reversal windows. Nevertheless, the intelligibility decrease showed the same pattern, namely, an initial decline which plateaued at 50% accuracy level, for both groups.

The experiment

In this study we use the locally time-reversed speech paradigm to investigate how our cognitive system temporally processes the incoming linguistic signal and how temporal distortion affects sign language comprehension. We are interested in disentangling the impact that signal modality and linguistic features may play in language decoding. The design had three separate tasks, each with a different type of material: Spanish sentences, LSE sentences and non-linguistic visual stimuli. This design lends itself to making two types of comparisons. Firstly, a comparison between spoken and signed language, in this case Spanish and LSE, can reveal how acoustic and visual perceptual properties modulate the temporal processing of language. Secondly, we want to compare sign language with non-linguistic visual material to investigate to what extent the results found in LSE are due to specific linguistic properties or general principles of the visual modality.

We selected non-linguistic stimuli (see Sections "Material" for detailed explanations of the stimuli) that were as comparable as possible to language in terms of regularity in their temporal structure; importantly, the non-linguistic stimuli had no phonological or syntactic structure. The same manipulation of local time reversal was applied to all three stimulus types, but the paradigm was slightly adapted according to the stimulus type. Reversal windows were different depending on the modality, and were chosen based on the results of previous studies employing this paradigm in spoken (Greenberg & Arai, 2001; Kiss, Cristescu, Fink, & Wittmann, 2008; Ueda et al., 2017) and signed (Hwang, 2011) languages. The range of reversal window sizes was balanced to detect both the intelligibility threshold and possible plateau effects at the extremes of the intelligibility curve. In line with the differences in temporal resolution between the auditory and visual modalities described above, reversal windows in the visual modality were larger and increased in larger steps than those used for spoken language.

If language information is temporally decoded in order to allow comprehension, and all types of linguistic signal rely on the same underlying temporal structure, we would expect to find similar patterns for both Spanish and LSE: a marked breakdown in intelligibility at a specific threshold. However, given the differences in the pseudo-periodicity of each type of signal (speech is a relatively fast changing signal compared to sign language) the threshold for intelligibility, which is typically around 40 ms for speech, should be at windows of a much longer duration for sign language. Conversely, different modalities may exploit temporal structure in qualitatively different ways: in this case, we expect Spanish and LSE to be differently affected by the temporal manipulation. The previous results found by Hwang (Hwang, 2011) suggest that this is the case. If signed languages are indeed characterized by a common temporal structure, and furthermore this temporal organization differs from that of spoken languages, we should find similar results when

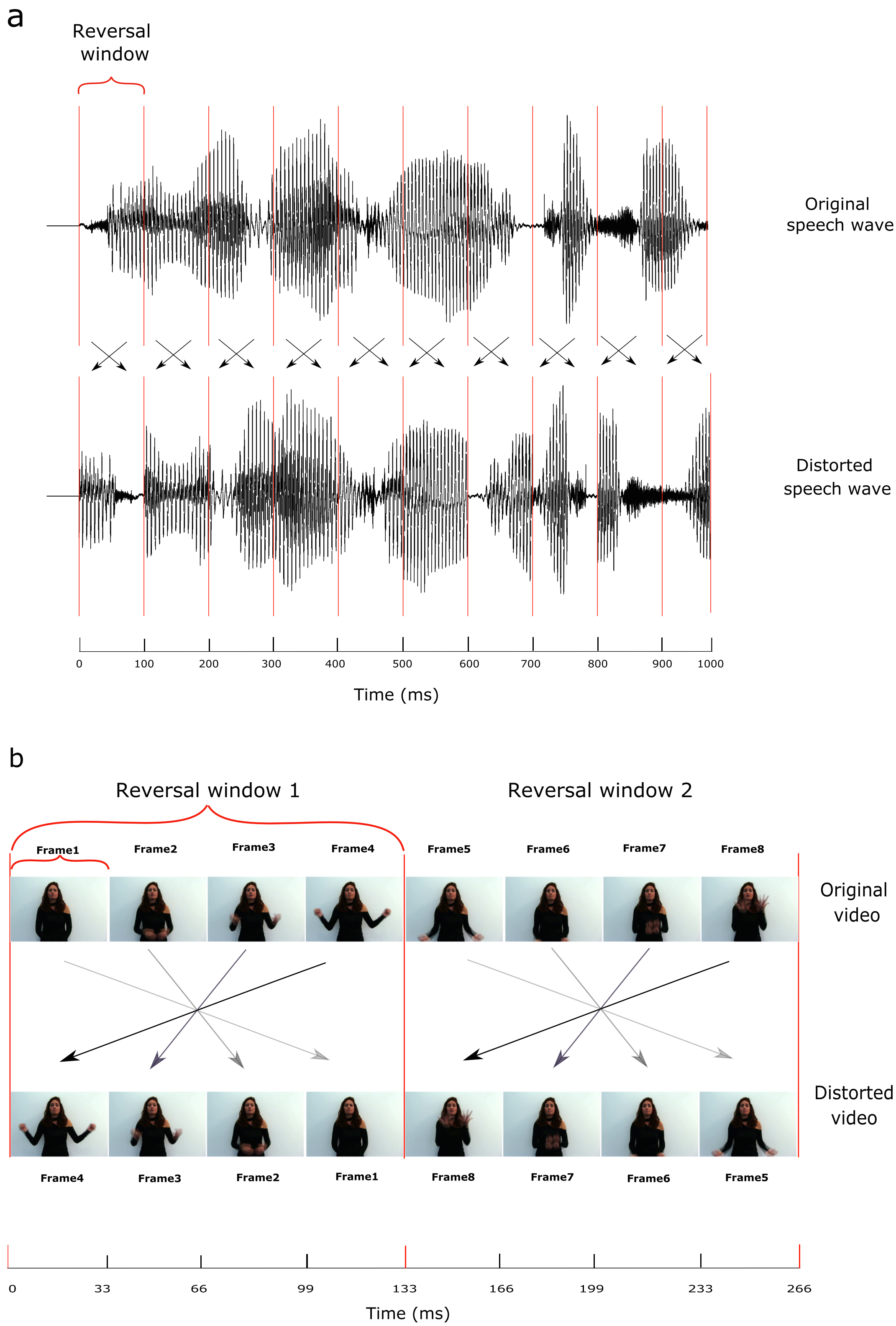

Fig. 1. A schematic to illustrate locally time-reversed speech paradigm applied to (a) Spanish and (b) LSE. In Spanish (a) the acoustic signal from the sentence 'El premio normal bloquea el tiempo preciso' ('The normal prize blocks the precise time') is divided into reversal windows of 100 ms and the signal is reversed inside each window, while the order of the windows themselves is maintained. In LSE (b) the video is made up of frames that last 33 ms each; the video is divided in reversal windows of 133 ms (4 frames that last 33 ms each), and the order of the frames is reversed within each window. The schematic shows the frames of the first sign, MILAGRO ['miracle'], of a sentence.

testing another sign language: LSE (Spanish Sign Language). Finally, if temporal structure is easier to segment when the input is linguistic, we should find a better performance in LSE. In contrast, if the visual signal is parsed in the same way and information extracted similarly when processing visual material, we should not find differences when comparing the LSE with a non-linguistic task.

The study consisted of a single session with three experimental tasks: Spanish, LSE and visual non-linguistic. The general characteristics of the study, with separate sections for each task to describe the materials and procedure, are presented in the following section (Section "Methods"). Afterwards we present the results of each task analysis and comparisons across different tasks (Section "Results") and we discuss the findings (Section "General discussion").

Methods

The experiment was conducted in different locations across Spain with the same equipment. The experiment ran on a DELL portable computer with Windows 7 OS, using Psychopy (version 1.85.3) in Python (version 2.7.11). All participants heard acoustic stimuli through headphones at the same comfortable volume. Responses in Spanish and LSE were recorded with a video camera. The order of the tasks was randomized across participants. All participants signed an informed consent form before the beginning of the experiment and were compensated for their participation.

The research was conducted with prior approval of BCBL Ethics Review Board and complied with the guidelines of the Helsinki Declaration.

Task 1: Spanish

Participants

Twenty-three participants took part in the experiment, 18 females and 5 males, with a mean age of 40 (29–51 years). All participants were bimodal bilinguals who were native speakers of Spanish and native or highly proficient users of LSE. Proficiency in LSE was reflected in self-reported ratings on a likert scale from 1 (no knowledge) to 5 (very good knowledge), with a mean rating of 4.81 (SD = 0.40) across all participants. Four participants were native signers, as they learned LSE before 1 years of age from a family member; those participants who were not native LSE signers were professional sign language interpreters.

Material

Stimuli consisted of 60 semantically unpredictable sentences in European Spanish, which were syntactically and grammatically correct sentences with no sensible meaning. The use of semantically unpredictable sentences assures that the results are due to correct perception of the sound and not to inference based on pragmatics or linguistic context (Greenberg & Arai, 2001; Hwang, 2011). We generated the sentences with a set of 120 adjectives, 120 nouns and 60 verbs. We selected the words from the European Spanish subtitle corpus of the EsPal database (Duchon, Perea, Sebastián-Gallés, Martí, & Carreiras, 2013) controlling for frequency (log_count between 3 and 7), number of phonemes per word (between 3 and 10) and number of syllables per word (between 2 and 4). Because the same words were used to generate the Spanish and LSE sentences (see Section "Material"), words that were homonyms in LSE were excluded, as were those which corresponded to compound signs or classifier-based signs with a very generic meaning.

The words were then randomly combined together to create 60 sentences with the following structure:

$Determiner_1 + Noun_1 + Adjective_1 + Verb + Determiner_2 + Noun_2 + Adjective_2$.

See Table 1 for example sentences. No noun, adjective or verb was repeated across sentences. A native speaker reviewed all the sentences to make sure that they were grammatically correct.

Table 1
Examples of semantically unpredictable sentences used in the experiment.

Spanish sentences	English translation
La luna urgente persigue al diablo bueno.	The urgent moon chases the good devil.
El interés ácido despide al cuchillo cojo.	The acid interest dismisses the lame knife.
El pueblo invisible divide el monstruo blanco.	The invisible village divides the white monster.
El campamento caro aguanta el puente azul.	The expensive camp holds the blue bridge.
El teatro contrario adora la distancia furiosa.	The opposite theatre adores the furious distance.
El empleo serio contiene la habilidad oscura.	The serious job contains the dark skill.
La mitad directa solicita la pasta negra.	The direct half requests the black pasta.
La planta favorita revisa la sangre fría.	The favourite plant checks the cold blood.
El vestido curioso demuestra el cuento vago.	The curious dress shows the vague tale.

A female native speaker recorded 65 sentences (60 experimental sentences and 5 practice sentences) with natural prosody and pace, using a Sennheiser ME65 microphone in a sound-proof recording booth. The item duration varied between 2.29 and 2.98 s (mean = 2.65 s, SD = 0.16). The recordings were normalized for sound level (70 dB) and distorted with the reversal window manipulation using Praat (Boersma, P., & Weenink, 2020). Five reversal windows were applied: 40 ms, 55 ms, 70 ms, 85 ms and 100 ms. At the beginning of each distorted sentence 50 ms of silence was added to avoid clipping due to the loading time of the audio file in the experiment presentation software.

Procedure

Participants heard 60 items, each in one of six conditions: undistorted or distorted with one of the five reversal windows. Items were pseudo-randomly assigned to different conditions across participants such that each participant heard ten items in each condition, and across all participants each item appeared the same number of times in each condition.

Participants read instructions in Spanish and then performed five practice items before beginning with the experimental trials. After hearing an item, participants could choose to respond or to hear the same item again three more times (up to a total of four presentations for the same item) before giving their answer (following Greenberg & Arai, 2001; Hwang, 2011). Participants gave their response by repeating out loud the sentence they believed that had heard. Responses were scored for word identity and word order. For word identity participants received one point for each lexical word that was correctly produced, up to a maximum of five points (given that there were five open class words in each sentence). For word order participants received up to four points for the correct relative order between each pair of lexical words in the sentence. The maximum score for each sentence was of nine points.

Task 2: Spanish sign language

Participants

Participants were the same as those for Task 1 (see Section "Participants").

Material

The material consisted of 60 semantically unpredictable sentences in Spanish Sign Language (LSE). We generated the sentences with the same sets of words used for the Spanish sentences (see Section "Material") by recombining the words together with the same random procedure. Each word was translated into the corresponding LSE sign taken from the Standard LSE Dictionary (Fundación CNSE, 2008) to avoid regional variants. All the signs had unique forms and none was a compound (to avoid phonological complexity) or a classifier-based sign with a generic meaning (to avoid semantic ambiguity). It was not possible to control for sign frequency and other lexical properties because this information is not available for LSE.

Following the grammar and syntax of LSE (Herrero Blanco, 2009) all sentences were recorded with the following SOV structure:

$Noun_1 + Adjective_1 + Noun_2 + Adjective_2 + Verb$.

No determiner sign was present in the sentences since LSE marks this feature by other means.

A deaf female native signer modeled 65 sentences in LSE (60 experimental sentences and 5 practice sentences) with natural prosody and pace, recorded at 25 fps with a video camera (Sony HDR-CX240E). The model signed in front of a uniform white background and each sentence began and ended with hands in a resting position in front of the body.

The videos of LSE sentences were pre-processed with FFmpeg (Tomar, 2006) (version 2.7): each video was cropped (810 × 540 pixels), the frame rate was set to 30 fps to match the 60 Hz refresh rate of the presentation screen, and the luminance was normalized across videos. The item duration varied between 3.36 and 5.76 s (mean = 4.67 s, SD = 0.48). The reversal-window manipulation was applied using a custom Python script and a fade in/out of 3 frames (100 ms) was applied to each of the manipulated sentences using FFmpeg. The reversal windows size for LSE sentences were: 4 frames (133 ms), 6 frames (199 ms), 8 frames (266 ms), 10 frames (333 ms) and 12 frames (399 ms).

Procedure

The procedure was the same as for Task 1, except that the participants saw the LSE stimuli on screen and gave their responses by signing.

Task 3: Visual non-linguistic stimuli

Participants

Participants were the same as those for Task 1 (see Section "Participants").

Material

The material consisted of 36 videos (30 experimental videos and 6 practice videos) representing four symbols traced one after the other in the center of the screen by a moving dot that left no line. The symbols used were six digits (1, 2, 3, 6, 8, 9) and six letters (C, L, O, S, V, Z). We chose to use letters and digits as symbols because they represent very well-known shapes that can be identified with a button press on a keyboard. Even if letters and digits have linguistic labels that may have been used to encode the elements in memory, the signal that presented the sequence is not language-like: the elements had no structure either internally (phonology) or in relation to each other (syntax). The pattern of the dot tracing the symbols was modeled after natural handwriting using a custom Matlab (version 2018b) script; all the symbols were written in one continuous stroke and were easily recognizable.

Each video consisted of a fixed sequence of two letters followed by two digits. The pool of 12 symbols was randomly combined together to fit this sequence, so that a symbol never appeared twice in a given trial and each symbol appeared the same number of times across all stimuli.

All videos were then manipulated using a custom Python script with the same procedure and the same reversal windows used for the sign language sentences: 4 frames (133 ms), 6 frames (199 ms), 8 frames (266 ms), 10 frames (333 ms) and 12 frames (399 ms)

Procedure

The procedure was similar to the one used in Task 2. Each participant saw each of the 30 videos in one of the six conditions (undistorted or with one of the five reversal windows). The condition in which a given item appeared was counterbalanced across participants. In contrast with the linguistic tasks, here participants could see each video only once before responding by typing the symbols they recognized on the keyboard. The difference was implemented to avoid ceiling effects; the results of a pilot session revealed high accuracy rates in Task 3 due in large part to the small set of stimuli used in the task. The instructions informed participants that the sequences consisted of two letters followed by two digits. Responses were scored by symbol identity (up to four points, one for each symbol) and symbol order (up to two points for the relative position between the two letters and the two digits), giving a maximum score of six points for each item.

Results

All statistical analyses for the experimental tasks were run using R (R Core Team, 2017) (version 3.6.2); mixed linear models were run using the lme4 package (Bates, Mächler, Bolker, & Walker, 2015) and analyzed with the lmerTest package (Kuznetsova, A., Brockhoff, P.B., & Christensen, 2013).

Analysis of sentence repetition

In both linguistic tasks participants could decide to hear or view each sentence up to four times before reproducing it. The distribution of the number of sentence presentations across reversal windows is different for Spanish and LSE, as shown in Fig. 2. While in Spanish participants requested more presentation of sentences with longer reversal windows, in LSE participants tended to view each sentence multiple times even when the reversal window was very short or in the (undistorted) baseline condition (we present a possible explanation for this strategy in Section "Comparison between spoken and sign language").

In order to investigate whether the number of sentence presentations modulates the intelligibility of the sentence we ran two separate linear mixed models for each language. Accuracy represented the dependent variable, while Reversal window and Number of presentations were input as continuous predictors; the random effects were intercepts for participant and item and a by-participant random slope for the effect of reversal window. The Spanish model ($R^2 = .77$) showed a statistically significant effect for reversal window ($\beta = -14.22$, $SE = 1.77$, $p < .001$), number of presentations ($\beta = -6.22$, $SE = 0.77$, $p < .001$) and the interaction between the two ($\beta = -8.17$, $SE = 0.6337$, $p < .001$). For Spanish sentences, intelligibility decreases with longer reversal windows and with a higher number of presentations. Lower accuracy scores were associated with longer reversal windows and, counterintuitively, with a higher number of sentence presentations: highly distorted sentences were more difficult to understand and therefore participants tended to hear them more times. The interaction effect is driven by the fact that a higher number of sentence presentations were associated with better performance only in the longer reversal windows: in shorter reversal windows, the pattern was reversed, as can be seen in Fig. 3.

The LSE model ($R^2 = .49$) showed a statistical significant effect for reversal windows ($\beta = -11.52$, $SE = 0.70$, $p < .001$), but the main effect of number of presentations ($\beta = -0.65$, $SE = 0.78$, $p = 0.39$) and the interaction ($\beta = 0.55$, $SE = 0.65$, $p = 0.39$) were not significant. In LSE participants viewed each sentence multiple times independently of the level of temporal distortion.

Number of sentence presentations was not included in the following analysis, where intelligibility is compared across different tasks.

Comparison between Spanish and LSE

To compare the results in LSE and Spanish we treated reversal windows in both languages as a categorical variable, making it possible to match the different absolute values of the windows used in each language (for a justification of reversal windows sizes see Section "The experiment"). We ran a linear mixed model including Reversal window and Task as categorical predictors; intercepts for participants and items were input as random effects. Main effects and interactions were assessed by calculating Type III F-statistics and significance p-values using Satterthwaite approximations to denominator degrees of freedom (Casaponsa et al., 2019). The model ($R^2 = .70$) showed a statistically significant effect for both Reversal window ($F_{(5,\ 2658)} = 792.02$, $p <$

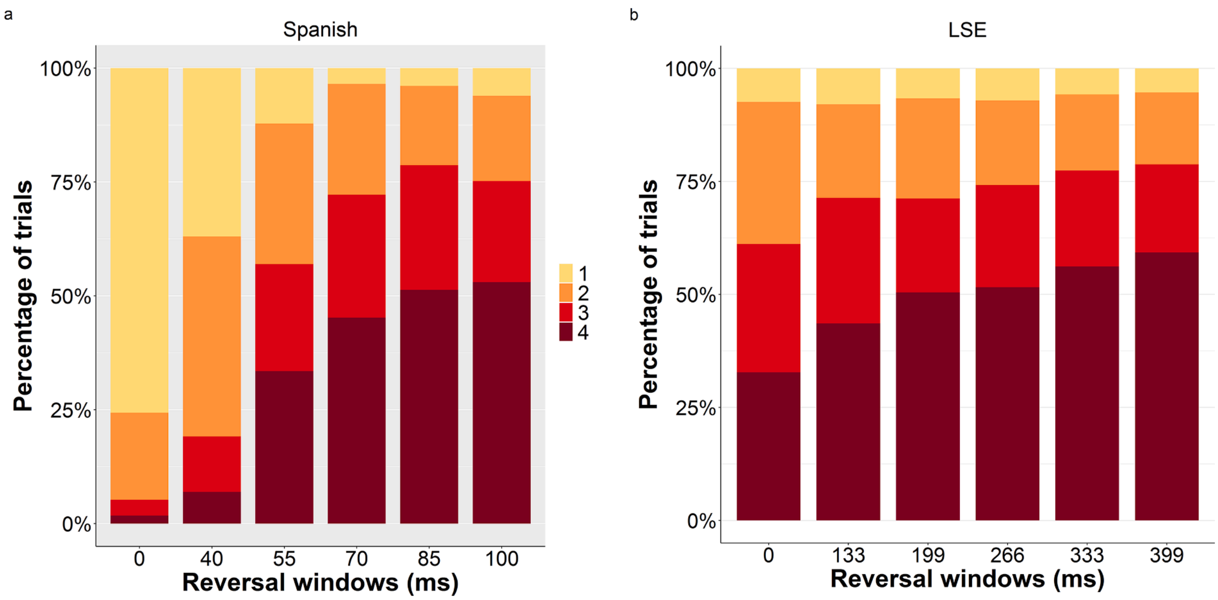

Fig. 2. Distribution of the number of stimulus presentations across reversal windows in (a) Spanish and (b) LSE.

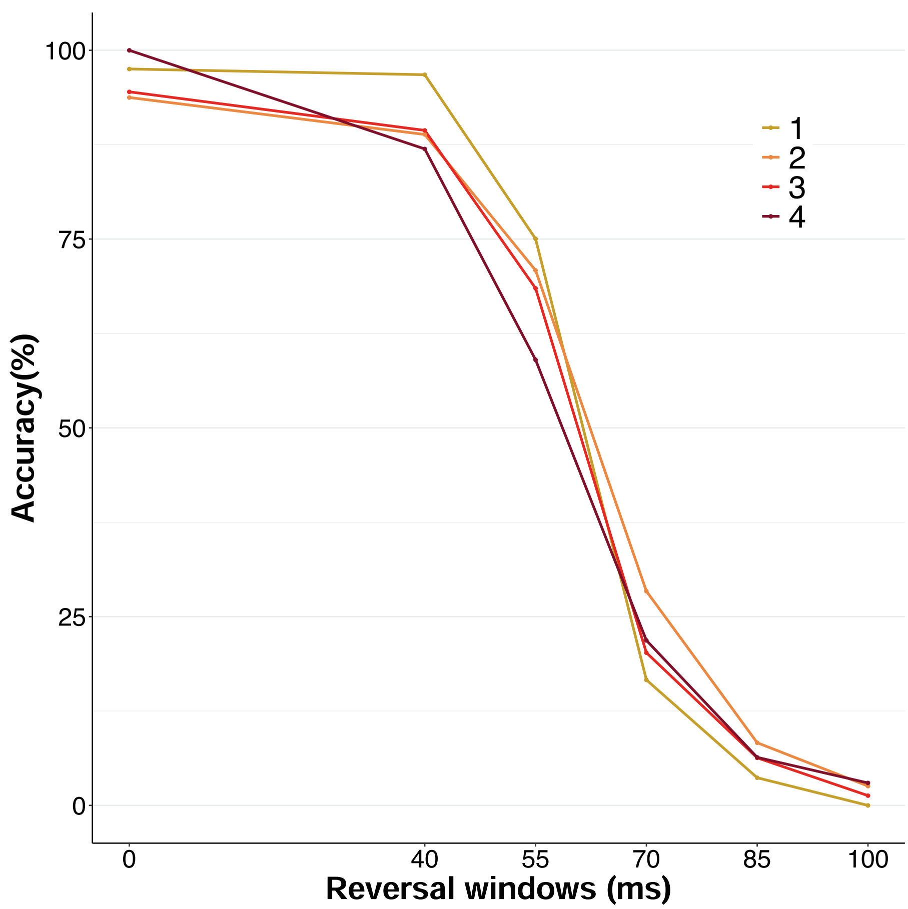

Fig. 3. Accuracy as a function of reversal window size and number of sentence presentations in Spanish.

.001) and Task ($F_{(1,\ 2657.1)} = 441.48$, $p < .001$). Moreover Reversal window and Task showed a significant interaction in modulating intelligibility ($F_{(5,\ 2658)} = 265.61$, $p < .001$). We performed a post hoc analysis on the model fitted values, comparing all consecutive reversal windows (rev1-rev2, rev2-rev3, rev3-rev4, rev4-rev5 and rev5-rev6) in each task and paired reversal windows across tasks. Consecutive post-hoc comparisons, corrected using the Bonferroni method, are shown in Table 2. Post-hoc comparisons across tasks are shown in Fig. 4 (for detailed values of the post-hoc analysis see Table S1 in the Supplementary Material).

Table 2
Summary of results of post-hoc t-tests (corrected using the Bonferroni method) comparing intelligibility between consecutive reversal windows for the Spanish and LSE stimuli. Post-hoc and descriptive statistics (mean and SD) are based on fitted data from the linear mixed model comparing the tasks in Spanish and LSE.

Reversal window	Spanish					*LSE*				
	Reversal window duration (ms)	Accuracy (%)		Comparison (with next window)		Reversal window duration (ms)	Accuracy (%)		Comparison (with next window)	
		Mean	SD	t	p		Mean	SD	t	p
1	0	96.9	2.26	2.67	0.12	0	86.6	2.27	9.45	<.001
2	40	91.7	2.26	12.85	<.001	133	68.3	2.27	0.62	1.00
3	55	66.8	2.26	22.77	<.001	199	67.1	2.27	3.43	0.01
4	70	22.7	2.26	8.39	<.001	266	60.3	2.27	3.28	0.02
5	85	6.46	2.26	2.16	0.48	333	53.9	2.27	1.66	1.00
6	100	2.27	2.26			399	50.6	2.27		

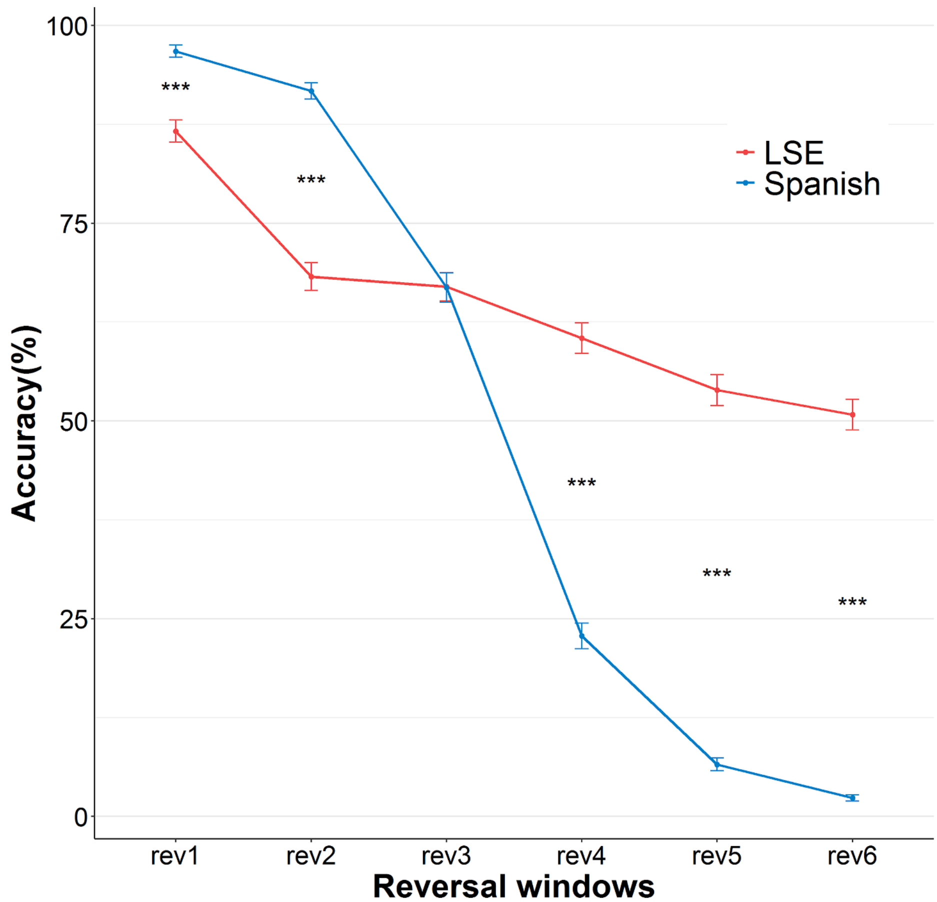

Fig. 4. Accuracy as a function of reversal window for the Spanish task (blue) and LSE task (red). Points joined by lines show intelligibility averaged across participants (n = 23) in each task, while error bars show the standard error of the mean for each point. Asterisks show statistically significant post-hoc comparisons across tasks in matched reversal windows (*** $p < .001$). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

In both tasks intelligibility decreased with increasingly larger reversal windows, but the pattern is different across the two languages. In Spanish temporal distortion with the smallest reversal window (40 ms) does not create any detrimental effect but intelligibility decreases sharply between window sizes of 40 and 85 ms, reaching an almost complete loss of word recognition. The results for LSE show a gradual decrease in intelligibility as the size of the reversal window increases: an initial dip between 0 and 133 ms followed by a less pronounced slope. For LSE, intelligibility scores spanned a limited range: in the baseline condition (0 ms) average intelligibility only reached 86% and even in the longest reversal window condition (399 ms) it never drops below 50%.

The comparison across paired reversal windows can inform us how increasing degrees of temporal distortion affect Spanish and LSE (Fig. 4). In the two smallest reversal windows intelligibility was higher in Spanish compared to LSE; in the three largest windows this pattern inverts as intelligibility in LSE was significantly higher than that of Spanish. Although intelligibility for undistorted sentences in LSE never reached 100%, these results reveal that sign language is more resilient than spoken language to this type of temporal manipulation: even under highly distorted conditions participants were able to recognize and repeat about half of the signs presented in the sentence. We return to these issues in the general discussion in Section “Comparison between spoken and sign language”.

Comparison between LSE and the visual non-linguistic task

To compare results in LSE and the visual non-linguistic task we carried out the same analysis we used to compare Spanish and LSE above. The model ($R^2 = .53$) showed a statistically significant effect for Reversal window ($F_{(5,\ 1965.9)} = 237.07$, $p < .001$), Task ($F_{(1,\ 2016,2)} = 190{,}57$, $p < .001$) and their interaction ($F_{(5,\ 1965.9)} = 33{,}48$, $p < .001$) in modulating intelligibility.

We performed post hoc analysis on the model fitted value, comparing all consecutive reversal windows (rev1-rev2, rev2-rev3, rev3-rev4, rev4-rev5 and rev5-rev6) in each task and paired reversal windows across tasks. Consecutive post-hoc comparisons, corrected with Bonferroni, are shown in Table 3; comparisons across tasks are shown in Fig. 5 (for detailed values of the post-hoc analysis see Table S2 in the Supplementary Material). A trend of lower accuracy with larger reversal windows was common to both tasks, with some differences.

The analysis of the visual non-linguistic task showed that accuracy in recognizing the symbols consistently lowered as the reversal window increases from 0 to 266 ms, but this drop in accuracy stopped at 333 ms. Participants found the task challenging, as reflected by the accuracy ranging between 87.7% and 19.2%. The comparison between performance in LSE and in the visual non-linguistic task (Fig. 5) evidences how accuracy is similar up until 133 ms reversal windows, but for larger windows the two curves start to diverge: the accuracy in LSE is higher than that of the visual non-linguistic task.

General discussion

The results from each task, taken individually, give us insight into how temporal distortion affects the perception of different types of signal. In Spanish it is possible to identify a clear threshold (40 ms reversal window) up to which incoming acoustic information is only minimally affected by temporal distortion. Once this threshold is reached though, the ability to perceive well-formed words from the acoustic signal is rapidly and almost completely lost. These results closely reproduce the findings in the literature on locally time-reversed speech: this pattern has been reported for various spoken languages (Ueda et al., 2017). The experiment extends the results of locally time-reversed speech paradigm to spoken Spanish, and adds to the growing body of evidence that the cognitive system perceives and parses different spoken languages with a common mechanism.

LSE, on the other hand, shows a more gradual pattern: longer reversal windows elicit an increasing loss of intelligibility, but the reduction is gradual and constant across reversal windows. Additionally, the cognitive system is still able to extract enough information to identify some of the signs in the distorted visual signal. The overall pattern in LSE matches the findings from the earlier study with deaf signers of ASL: in that study, participants' accuracy gradually decreased and spanned between 90% and 50% (Hwang, 2011). That study and the current experiment represent the only two studies applying locally time-reversed paradigm to signed languages; more research is therefore needed to draw strong conclusions but the similarities between the LSE and ASL results suggest that there may be a common temporal mechanism for visual signed language.

In the following sections we compare the results across language modality (spoken vs sign) and stimulus type (linguistic vs non-linguistic) to gain a fuller picture of the role played by modality and linguistic structure in temporal processing.

Comparison between spoken and sign language

The main goal of this study was to investigate how language modality impacts the way our cognitive system analyses and parses the incoming linguistic signal. Spanish and LSE differ qualitatively in how intelligibility is modulated by increasingly longer reversal windows. The difference between Spanish and LSE resides in the absence of a clear drop in intelligibility corresponding to a specific reversal window size in LSE. This reversal window, sometimes defined as temporal integration window (Poeppel, 2003), represents the temporal resolution unit for spoken language processing.

In general LSE appears to be more resistant to temporal distortion than Spanish. Two factors may explain the resilience of the sign language signal. Firstly, signed languages make use of the visual domain, which relies less on temporal structure. Intuitively, this makes sense: we have little difficulty in reversing a movement, but reversing a sound is a much more onerous task. Support also comes from neuroimaging studies that have used backwards signing as a baseline condition and report that signers could identify some lexical items (Inubushi & Sakai, 2013). In a recent study Bosworth et al. (2020) presented ASL narratives backwards, and found that signers were still able to understand and recall a good part of the narrative. Secondly, the spatial character of sign languages means that information is maintained over time: phonological parameters, such as handshape, are stable in the signal for enough time to make it possible to recover these features even when the temporal properties are degraded. Evidence from the production of backwards signing supports this idea that groupings of phonological features remain accessible within the temporal syllabic structure of the sign (Wilbur & Petersen, 1997). Furthermore, the use of space allows different features to appear at the same time: identifying a combination of features, such as a handshape and a location, may be sufficient to identify a lexical item even when all the other parameters are unidentifiable or highly distorted. This relates to the distribution of information in the sign language lexicon and possible redundancies in the signal. This redundancy is evident also in previous work on the lexical recognition of signed and spoken languages: Emmorey and Corina (1990) found that lexical recognition is faster and easier for signs compared to spoken words and argued that this is due to spatial and temporal properties associated with sign structure.

Despite the resilience associated with signed languages, LSE performance in conditions with little or no distortion was lower than that of

Table 3
Summary of results of post-hoc t-tests (corrected using the Bonferroni method) comparing intelligibility between consecutive reversal windows for the visual non-linguistic stimuli and LSE stimuli. Post-hoc and descriptive statistics (mean and SD) are based on fitted data from the linear mixed model comparing the visual non-linguistic and LSE tasks.

Reversal window	Non-linguistic stimuli					*LSE*				
	Reversal window duration (ms)	Accuracy (%)		Comparison (with next window)		Reversal window duration (ms)	Accuracy (%)		Comparison (with next window)	
		Mean	SD	t	p		Mean	SD	t	p
1	0	87.7	3.39	6.03	<.001	0	86.6	3.01	8.72	<.001
2	133	69.8	3.39	5.85	<.001	133	68.3	3.01	0.52	1.00
3	199	52.4	3.39	5.71	<.001	199	67.2	3.01	3.24	0.02
4	266	35.5	3.39	3.82	<.01	266	60.3	3.01	2.95	0.05
5	333	24.1	3.39	1.64	1.00	333	54.1	3.01	1.66	1.00
6	399	19.2	3.39			399	50.5	3.01		

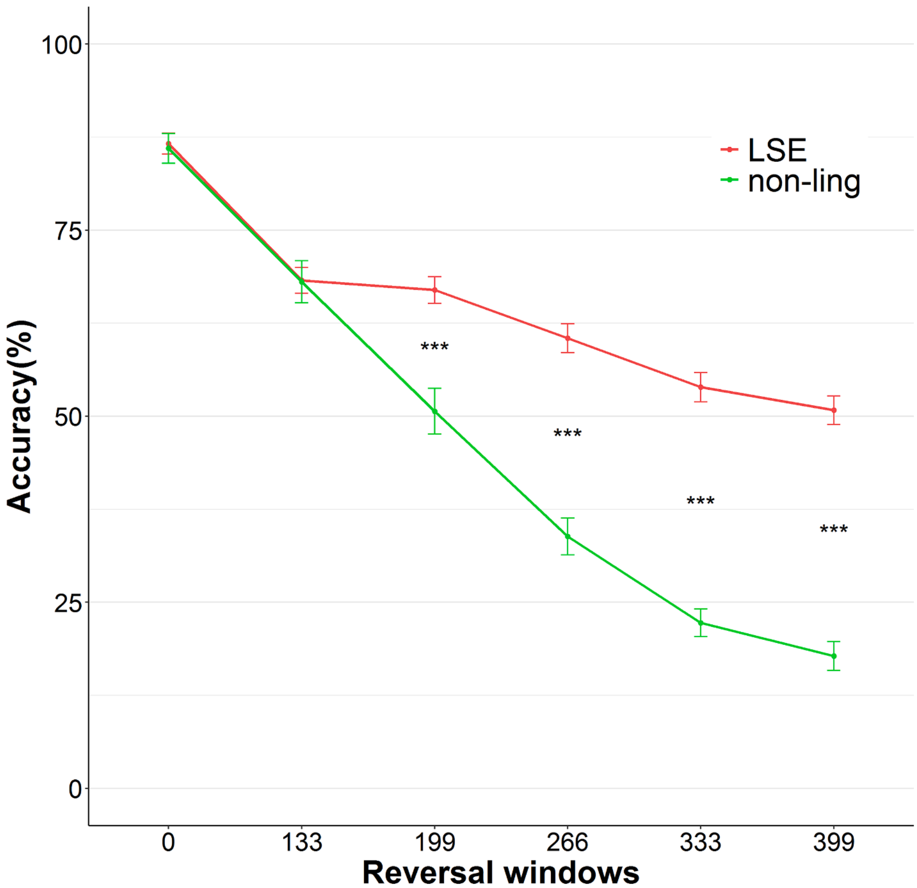

Fig. 5. Accuracy as a function of reversal window size the visual non-linguistic task (green) and sign language task (red). Points joined by lines show intelligibility averaged across participants (n = 23) in each task, while error bars show the standard error of the mean for each point. Asterisks show statistically significant post-hoc comparisons across tasks in matched reversal windows (*** p < .001). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Spanish, and never reached ceiling accuracy. An explanation for this result might lie in differences in task difficulty. Participants were all native Spanish speakers, whereas for most of them LSE was their second language (albeit with a high level of proficiency). Language proficiency of participants might therefore account for the lower performance in undistorted sentences and the higher level of variability in participants' performance during the task in sign language. Age of acquisition does modulate performance on this type of task: when the locally time-reversed paradigm is performed by non-native speakers intelligibility starts to decline at shorter reversal windows compared to native speakers, but the general trend and shape of the intelligibility curve is similar across native and non-native participants (for spoken German, (Kiss et al., 2008); for ASL, (Hwang, 2011)). Being a non-native speaker (or signer) modulates resilience to locally time-reversed speech, but the overall pattern remains unchanged. Nevertheless, the same effect is evident in the previous ASL study (Hwang, 2011), in which deaf native signers had a baseline accuracy of approximately 90%, comparable to our result. Another independent study found similar results while testing the intelligibility of time-compressed ASL sentences in deaf native signers: in this population the accuracy for sentences played at the normal rate was at 88% (Fischer et al., 1999). The errors for undistorted LSE (and ASL) can be explained by the higher short term memory effort associated with signed languages. Short term memory span in signed language is about 5 ± 2, compared to the classical 7 ± 2 associated with auditory presented stimuli (Boutla, Supalla, Newport, & Bavelier, 2004). Different explanations have been proposed to account for this difference, such as a phonological similarity effect across signs or sign articulatory length; recent work with second language learners of signs suggests a possible role of perceptual-motor memory processes in the sign lexicon (Martinez & Singleton, 2018). The lower baseline accuracy for sign languages appears to be related not to age of acquisition but to linguistic properties that affect processing; future work comparing native and non-native signers could tease apart the relative contribution of proficiency, age of acquisition and modality.

In both linguistic tasks participants could hear or see the same sentence up to four times before giving their answer. In the spoken language task the interaction of number of presentations and reversal windows size modulated intelligibility. More repetitions improved performance in the longer reversal windows (70–100 ms), while in shorter reversal windows fewer repetitions were associated with better intelligibility. In the sign language task, on the other hand, we observed a different distribution in the number of repetitions: most participants tended to watch each sentence three or four times, independently of the size of the reversal window. Participants reported struggling to retain in memory the signs from the sentence much more than they did with Spanish words; as a result they often decided to see the sentence more than once, even in the baseline condition. This strategy seems to be driven by the greater short term memory effort associated with maintaining and retrieving signed language items (Boutla et al., 2004; Hall & Bavelier, 2011). Since nearly all of our participants were not native signers, non-native processing might be driving this effect. However, data from a sentence repetition task in ASL show that (deaf and hearing) native signers also struggle with this sort of task, and, more importantly, suggest that fluency (rather than hearing status) is what modulates the type of working memory strategy employed to carry out the task (Supalla, Hauser & Bavelier, 2014). The number of repetitions of sign language sentences in our study does not modulate accuracy because participants chose to see most items as many times as possible.

Overall our results indicate that sign language does not share the same temporal resolution as spoken language, and suggest that the temporal processing of language relies on a mechanism which is at least partly modality dependent. We acknowledge some limitations in the comparison between Spanish and LSE. The manipulation applied to spoken and signed language may affect the acoustic and the visual modality differently. The locally time-reversed speech paradigm is designed to distort the temporal order of the signal. As the spatial domain plays a fundamental role in signed languages, a distortion that specifically targets its spatial organization — as opposed to the temporal one — might impair language intelligibility to greater extent. Moreover, age of acquisition and proficiency level of LSE could partly modulate their performance in the task, although our results (with hearing bimodal bilinguals) concur with those of deaf native signers (Hwang, 2011). Nevertheless, future studies employing spatial distortions and investigating a different population (native signers and less proficient signers) could contribute to a better understanding of sign language structure and processing.

Comparison between linguistic and non-linguistic material

Comparing results from the LSE task and the visual non-linguistic task casts light on the specific role of language properties in the temporal parsing of a visual signal. In LSE participants had to reproduce signs within a sentence, while in the non-linguistic task they saw videos of a dot tracing a sequence of four easily recognizable and nameable symbols (two letters and two digits). The sequential presentation of the symbols, and the fact that participants had to identify the symbols in the correct order, parallels the recognition of signs in the LSE task. Overall, we believe that our non-linguistic stimuli can provide a valid term of comparison, although the tasks are not identical. An important difference is the structure of the stimuli: while both types of materials rely on the spatial dimension, the multiple articulators which characterize LSE are not present in the non-linguistic task, which means that information is not presented simultaneously. Moreover, the sets of stimuli used in the two tasks are different: 300 signs in LSE compared to only 12 symbols for the non-linguistic task. The probability of presentation for each symbol is therefore much higher than that for the signs. Another difference lies in the paradigm used in the two tasks: while in LSE participants could view each sentence up to four times before giving their answer, in the visual non-linguistic task they could see each video only once. The analysis showed that the number of sentence repetitions in the LSE task had no significant effect on language intelligibility but this null effect may be due to the fact that participants chose to view most sentences four times, regardless of the degree of distortion. More critically, the two tasks showed the same levels of accuracy for the undistorted stimuli, as can be seen in Fig. 5, suggesting that in the baseline condition they were comparable. The accuracy curves start to diverge only after the second reversal window and this difference between the two tasks cannot be ascribed to a simple difference in how many times the stimuli were viewed by participants.

The results show that in both tasks the accuracy in perceiving distorted stimuli gradually decreases as reversal window size increases. In contrast to the response for spoken language, neither curve has a clear point where intelligibility and recognition of the stimuli is lost, suggesting that the visual domain is not as susceptible to temporal disturbance. Although the pattern is very similar across the visual tasks, in the non-linguistic task accuracy drops more than in the LSE task: intelligibility in LSE never drops under 50% while in the non-linguistic task accuracy falls to 20%. This difference is surprising, since intuitively symbol recognition should be easier than sign recognition given the restricted set of symbols used in the experiment. Both tasks benefit from the superiority of the visual modality in managing temporal distortion, but LSE shows some additional advantage. As mentioned in the comparison of the spoken and sign language results (Section “Comparison between spoken and sign language”), the spatial and temporal structure of signs makes possible an over-representation of the linguistic information in the signal. Evidence comes from studies where participants were still able to recognize signs with high accuracy even when the information in the signal was reduced, for example when the videos were presented with fewer frames per second (Johnson & Caird, 1996) or speeded up by a factor of three (Fischer, Delhorne, & Reed, 1999). The relative resilience to temporal distortion of LSE with respect to the non-linguistic stimuli suggests that features of the linguistic signal in this modality aid recognition of temporally distorted signs. These features may include the simultaneous articulation afforded by the visuo-spatial channel as well as the combinatorial properties of sub-lexical units of sign language.

The results point toward a common mechanism for the temporal resolution of visually-presented stimuli that is characterized by a constant reduction in accuracy as the signal becomes more distorted temporally. The organization of visual information in signed languages appears to attenuate how much information is lost by any disturbance of the temporal structure.

Conclusion

The fundamental role that temporal structure plays in speech comprehension, calls for a clear characterization of the temporal properties of language processing more generally. Spoken and signed languages make use of two different sensory channels (the acoustic and the visual channel); their comparative study represents a unique opportunity to investigate to what degree modality shapes the temporal structure of language. Our results suggest that temporal language processing arises from the interaction between the properties of the sensory system and the special characteristics of language. The perceptual modality poses constraints on how linguistic information is optimally processed by our cognitive system: the visual and auditory systems are characterized by different properties and perceptual processing of the physical signal will be different. At the same time, language structure accommodates the advantages and limits of a specific modality. This is particularly clear in the case of spoken and signed languages, where different temporal structures reflect the sequential or parallel organization of the information. Within the visual modality the results for the language signal shows an advantage compared to non-linguistic material, suggesting the informational and temporal properties of the language signal favor its processing by our cognitive system. The reciprocal influence that language and sensory modality plays in shaping the temporal structure of the signal is extremely complex, and this study provides a first step towards disentangling their specific contributions.

The data, analysis scripts and example stimuli are available in the following repository: https://osf.io/qr38u/?view_only=56e75d3c9c5e461eb03518124dd301e1.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We are grateful to the organizations that provided the premises to run the experiment: ASORNA (Deaf Association in Navarra), EUNATE (Association of Families of Deaf People in Navarra), Castaños Civic Center (Bilbao), The University of Valladolid and the López Vicuña Vocational Institute (Palencia). We would like to thank Saul Villameriel, Paco Vera Villaverde and Aina Casaponsa, who provided support during data collection, processing and analysis. Finally, we are truly indebted to all our 23 participants.

The authors acknowledge financial support from the Basque Government through the BERC 2018-2021 program, from the Spanish Ministry of Science and Innovation (Agencia Estatal de Investigación, AEI), through grants (PSI-2016-76435-P to B.C.; and RTI2018-093547-B-I00 to M.C.), through Juan de la Cierva Fellowship (FJCI-2017-31806 to B.C.) and from “La Caixa” Foundation and the European Union’s Horizon 2020 research and innovation programme under the Marie-Skłodowska-Curie grant agreement no. 713673 (personal fellowship code LCF/BQ/DI18/11660008 to C.L.R.). This research was founded by the Severo Ochoa Center of Excellence grant (CEX2020-001010-S).

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jml.2021.104283.

References

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1). https://doi.org/10.18637/jss.v067.i01.

Bellugi, U., & Fischer, S. (1972). A comparison of sign language and spoken language. *Cognition, 1*(2–3), 173–200. https://doi.org/10.1016/0010-0277(72)90018-2.

Boersma, P., & Weenink, D. (2020). Praat: doing phonetics by computer [Computer program] Version 6.1. 16. Online at http://www.praat.Org.

Bosworth, R., Stone, A., & Hwang, S. O. (2020). Effects of video reversal on gaze patterns during signed narrative comprehension. *The Journal of Deaf Studies and Deaf Education, 25*(3), 283–297.

Boutla, M., Supalla, T., Newport, E. L., & Bavelier, D. (2004). Short-term memory span: Insights from sign language. *Nature Neuroscience, 7*(9), 997–1002. https://doi.org/10.1038/nn1298.

Brentari, D. (1998). *A prosodic model of sign language phonology*. Mit Press.

Casaponsa, A., Sohoglu, E., Moore, D. R., Füllgrabe, C., Molloy, K., Amitay, S., & Malmierca, M. S. (2019). Does training with amplitude modulated tones affect tone-vocoded speech perception? *PLoS ONE, 14*(12), e0226288. https://doi.org/10.1371/journal.pone.0226288.

Cummins, F. (2012). Oscillators and syllables: A cautionary note. *Frontiers in Psychology, 3*(OCT). https://doi.org/10.3389/fpsyg.2012.00364.

Ding, N., Melloni, L., Yang, A., Wang, Y., Zhang, W., & Poeppel, D. (2017). Characterizing neural entrainment to hierarchical linguistic units using electroencephalography (EEG). *Frontiers in Human Neuroscience, 11*(September), 1–9. https://doi.org/10.3389/fnhum.2017.00481.

Doelling, K., Arnal, L., Ghitza, O., & Poeppel, D. (2014). Acoustic landmarks drive delta-theta oscillations to enable speech comprehension by facilitating perceptual parsing. *Neuroimage, 85*(15). https://doi.org/10.1016/j.neuroimage.2013.06.035.Acoustic.

Duchon, A., Perea, M., Sebastián-Gallés, N., Martí, A., & Carreiras, M. (2013). EsPal: One-stop shopping for Spanish word properties. *Behavior Research Methods, 45*(4), 1246–1258. https://doi.org/10.3758/s13428-013-0326-1.

Emmorey, K., & Corina, D. (1990). Lexical recognition in sign language: Effects of phonetic structure and morphology. *Perceptual and motor skills, 71*(3_suppl), 1227–1252. https://doi.org/10.2466/pms.1990.71.3f.1227.

Fischer, S. D., Delhorne, L. A., & Reed, C. M. (1999). Effects of rate of presentation on the reception of american sign language. *Journal of Speech, Language, and Hearing Research, 42*(3), 568–582. https://doi.org/10.1044/jslhr.4203.568.

Fundación CNSE (2008). Diccionario normativo de la lengua de signos española. Fundación CNSE.

Greenberg, S., & Arai, T. (2001). The relation between speech intelligibility and the complex modulation spectrum. Proceedings of the 7th European Conference on Speech Communication and Technology, 473–476. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.66.2050&rep=rep1&type=pdf.

Grosjean, F. (1977). The perception of rate in spoken and sign languages. *Perception & Psychophysics, 22*(4), 408–413. https://doi.org/10.3758/BF03199708.

Hall, M. L., & Bavelier, D. (2011). Short-term memory stages in sign vs. speech: The source of the serial span discrepancy. *Cognition, 120*(1), 54–66. https://doi.org/10.1016/j.cognition.2011.02.014.

Herrero Blanco, Á. L. (2009). Gramática didáctica de la lengua de signos española (LSE). Sm.

Hwang, S.-O.-K. (2011). Windows into sensory integration and rates in language processing: Insights from signed and spoken languages. *Dissertation Abstracts International Section A: Humanities and Social Sciences*.

Inubushi, T., & Sakai, K. L. (2013). Functional and anatomical correlates of word-, sentence-, and discourse-level integration in sign language. *Frontiers in Human Neuroscience, 7*(OCT), 681. https://doi.org/10.3389/fnhum.2013.00681.

Johnson, B. F., & Caird, J. K. (1996). Effect of frame rate and video information redundancy on the perceptual learning of American sign language gestures. *Conference on Human Factors in Computing Systems - Proceedings, 121–122*. https://doi.org/10.1145/257089.257200.

Kiss, M., Cristescu, T., Fink, M., & Wittmann, M. (2008). Auditory language comprehension of temporally reversed speech signals in native and non-native speakers. Retrieved from *Acta Neurobiologiae Experimentalis, 68*(2), 204–213 http://www.embase.com/search/results?subaction=viewrecord&from=export&id=L351871137.

Klima, E. S., & Bellugi, U. (1979). *The signs of language*. Harvard University Press.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2013). lmerTest: Tests for random and fixed effects for linear mixed effect models (lmer objects of lme4 package). *Version, 2*.

MacSweeney, M., Capek, C. M., Campbell, R., & Woll, B. (2008). The signing brain: The neurobiology of sign language. *Trends in Cognitive Sciences, 12*(11), 432–440. https://doi.org/10.1016/j.tics.2008.07.010.

Martinez, David, & Singleton, Jenny L. (2018). Predicting sign learning in hearing adults: The role of perceptual-motor (and phonological?) processes. *Applied Psycholinguistics, 39*(5), 905–931.

Meier, R. P. (2002). Why different, why the same? Explaining effects and non-effects of modality upon linguistic structure in sign and speech. In Modality and structure in signed and spoken languages (pp. 1–25).

Meyer, L. (2018). The neural oscillations of speech processing and language comprehension: State of the art and emerging mechanisms. *European Journal of Neuroscience, 48*(7), 2609–2621. https://doi.org/10.1111/ejn.13748.

Poeppel, D. (2003). The analysis of speech in different temporal integration windows: Cerebral lateralization as "asymmetric sampling in time". *Speech Communication, 41*(1), 245–255. https://doi.org/10.1016/S0167-6393(02)00107-3.

Poeppel, D., & Assaneo, M. F. (2020). Speech rhythms and their neural foundations. *Nature Reviews Neuroscience, 21*(6), 322–334. https://doi.org/10.1038/s41583-020-0304-4.

Samuel, A. G. (1991). Perceptual degradation due to signal alternation: Implications for auditory pattern processing. *Journal of Experimental Psychology: Human Perception and Performance, 17*(2), 392–403. https://doi.org/10.1037/0096-1523.17.2.392.

Samuel, A. G. (2020). Psycholinguists should resist the allure of linguistic units as perceptual units. *Journal of Memory and Language, 111*, 104070. https://doi.org/10.1016/j.jml.2019.104070.

Sandler, W. (2011). *Phonological representation of the sign: Linearity and nonlinearity in American Sign Language* (Vol. (Vol. 32). Gruyter., Walter de.

Sandler, W. (2018). The body as evidence for the nature of language. *Frontiers in Psychology, 9*(OCT), 1–21. https://doi.org/10.3389/fpsyg.2018.01782.

Steffen, A., & Werani, A. (1994). An experiment on temporal processing in language perception. *Sprechwissenschaft Und Psycholinguistik, 6*, 189–205.

Stilp, C. E., Kiefte, M., Alexander, J. M., & Kluender, K. R. (2010). Cochlea-scaled spectral entropy predicts rate-invariant intelligibility of temporally distorted sentencesa). *The Journal of the Acoustical Society of America, 128*(4), 2112–2126. https://doi.org/10.1121/1.3483719.

Stokoe, W. C., & Marschark, M. (2005). Sign language structure: An outline of the visual communication systems of the american deaf. *Journal of Deaf Studies and Deaf Education, 10*(1), 3–37. https://doi.org/10.1093/deafed/eni001.

Supalla, T., Hauser, P. C., & Bavelier, D. (2014). Reproducing American Sign Language sentences: Cognitive scaffolding in working memory. *Frontiers in Psychology, 5*, 859.

Tomar, S. (2006). Converting video formats with FFmpeg. *Linux Journal, 146*(10).

Ueda, K., Nakajima, Y., Ellermeier, W., & Kattner, F. (2017). Intelligibility of locally time-reversed speech: A multilingual comparison. *Scientific Reports, 7*(1), 1–8. https://doi.org/10.1038/s41598-017-01831-z.

Walsh, B., & Smith, A. (2002). Articulatory movements in adolescents: Evidence for protracted development of speech motor control processes. *Journal of Speech, Language, and Hearing Research, 45*(6), 1119–1133. https://doi.org/10.1044/1092-4388(2002/090).

Wilbur, R. B. (2009). Effects of varying rate of signing on ASL manual signs and nonmanual markers. *Language and Speech, 52*(2–3), 245–285. https://doi.org/10.1177/0023830909103174.

Wilbur, R. B., & Petersen, L. (1997). Backwards signing and ASL syllable structure. *Language and Speech, 40*(1), 63–90.

R Core Team (2017) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.