

Article

Not peer-reviewed version

Is Statistical Learning Affected by the Composition of the Speech Streams? Behavioral and Neural Evidence

[Ana Paula Soares](#)*, Dario Paiva, [Alberto Lema](#), [Diana R. Pereira](#), Ana Cláudia Rodrigues, [Helena Mendes Oliveira](#)

Posted Date: 3 January 2025

doi: 10.20944/preprints202501.0151.v1

Keywords: statistical learning; implicit learning; explicit learning; entropy; artificial languages



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Article

Is Statistical Learning Affected by the Composition of the Speech Streams? Behavioral and Neural Evidence

Ana Paula Soares ^{1,*}, Dario Paiva ¹, Alberto Lema ^{1,2}, Diana R. Pereira ¹, Ana Cláudia Rodrigues ^{1,2} and Helena Mendes Oliveira ¹

¹ Human Cognition Lab, CIPsi, School of Psychology, University of Minho, Portugal

² Psychological Neuroscience Lab, CIPsi, School of Psychology, University of Minho, Portugal

* Correspondence: asoares@psi.uminho.pt

Abstract: Statistical learning (SL), the ability to extract patterns from the environment, has been assumed to play a central role in whole cognition, particularly in language acquisition. Evidence has been gathered, however, from behavioral experiments relying on simplified artificial languages, raising doubts on the generalizability of these results to natural contexts. Here, we tested if SL is affected by the composition of the streams by expositing participants to auditory streams containing either four nonsense words presenting a transitional probability (TP) of 1.0 (unmixed high-TP condition), four nonsense words presenting TPs of .33 (unmixed low-TP condition) or two nonsense words presenting TP of 1.0 and two TP of .33 (mixed condition), first, under incidental (implicit), and, subsequently, under intentional (explicit) conditions to further ascertain how prior knowledge modulates the results. Electrophysiological and behavioral data were collected from the familiarization and test phases of each of the SL tasks. Behavior results revealed reliable signs of SL for all the streams, even though differences across stream conditions failed to reach significance. The neural results revealed, however, facilitative processing of the mixed over the unmixed low-TP and the unmixed high-TP conditions in the N400 and P200 components, suggesting that moderate levels of entropy boost SL.

Keywords: statistical learning; implicit learning; explicit learning; entropy; artificial languages

1. Introduction

Statistical learning (SL), the ability to extract patterns from the sensory environment even without intention or awareness, has been assumed to play a central role in whole cognition, particularly in the learning of the rule-governed aspects of language. The first evidence of this comes from a seminal work conducted by Saffran et al. [1] showing that eight-month-old babies exposed for two min to a continuous speech stream made of the repetition of three-syllable nonsense words (e.g., “gikoba”, “tokibu”, “tipolu”, “gopila”) presented with no pauses between each other and with no repetition of the same “word” in a row (e.g., “gikobatokibutipolugopilatokibu”), were able to extract the co-occurrences between adjacent syllables - a statistics known as transitional probability (TP), to extract word-like units (e.g., “gikoba”). Note that, in that artificial language, the probability of a syllable such as “ko” to be followed by “gi”, and a syllable such as “ba” to follow “ko” is highly likely, whereas the probability of a syllable such as “go” to succeed “ba” is less likely once each “word” could only be followed by another “word” in the stream with the same level of probability. Once TPs were the only cue available to assist “word” segmentation, the results obtained from this seminal study suggested that babies use the statistical regularities embedded in speech to discover “word” boundaries.

Since then, many other works have shown that SL mechanisms are also present in other levels of language acquisition, such as word-referent associations (e.g., [2–4]), grammatical categorization

(e.g., [5,6]), the establishment of long-distance dependencies (e.g., [7–10]), and the development of literacy skills (e.g., [11–14]). However, as Siegelman recently pointed out [15], even if all these studies provide evidence that SL is a powerful mechanism that enables individuals to crack the language code and acquire language so quickly and effortlessly, the fact that numerous works have shown that humans possess remarkable abilities to detect regularities in the environment and that language is a system full of regularities at multiple levels (phonological, orthographic, morphosyntactic, and grammatical), this does not necessarily mean that SL mechanisms are used and play a fundamental role in language acquisition.

Alternative approaches claim that SL mechanisms cannot fully account for the processes underlying language acquisition in “real” contexts, as these conclusions were drawn from laboratory studies using oversimplified artificial languages that do not mimic the complexity of natural languages (see [16] and [17] for discussion). Indeed, the vast majority of the SL studies conducted so far have exposed participants not only to a small number of nonsense words (typically four or six), which appear in the stream the same number of times to account for word frequency effects, but also “words” presenting TPs of 1.0, meaning that a given syllable only occurs in a given “word” at a specific syllable position (e.g., [1,18–22]). However, as Soares et al. recently emphasized [23], in natural languages, words show a much more diversified pattern both in terms of syllable length and syllable composition and also in the number of times each word occurs in the language, showing a Zipfian distribution (see [24,25] and also [26–28] for evidence in the European Portuguese language).

All these aspects considerably change the proprieties of the input to which participants are exposed in natural vs. artificial languages, casting doubts on the generalizability of the results obtained from lab experiments to “real” contexts (see [29] and also [30] for reviews). Furthermore, in these studies, it is frequently assumed that participants are *tabula rasae* with no prior knowledge about the language to which they have been exposed (see [31] and [32]), which does not also mimic the way language is acquired in natural contexts once it is well known that individuals use previous linguistic knowledge to assist the acquisition of the new one through anchoring and/or bootstrapping mechanisms (e.g., see [33] for details).

To address these issues, recent SL studies have attempted to analyze how SL works in a wide range of conditions (e.g., [19,20,34–38]). Of particular interest for this paper are the studies conducted by Soares et al. ([23,39–42]), with adult and children’s participants with and without developmental language disorder (DLD; previously named specific language impairment [SLI] - see [43] and [44] for details) using the triplet-embedded paradigm introduced by Saffran et al. [1] from which behavioral and neural data were collected. The behavioral task used however to test SL after the exposure phase was the two-alternative forced-choice (2-AFC) task in which participants were asked to decide which stimuli in a pair (a three-syllable nonsense word presented during familiarization and a foil made of the same syllables in an order not presented before) “sounded more familiar” and not by analyzing the times babies spent listening familiar (i.e., “words” presented before) vs. nonfamiliar “words” (i.e., “words” made of the same syllables presented in a new order) as in Saffran et al.’s [1] seminal study.

Specifically, in these studies, Soares and colleagues ([23,39–42]), exposed participants to speech streams made of the repetition of eight three-syllable nonsense words, four of which presenting TPs of 1.0 as in [1] (called *easy* or *high-TP words*), and four TPs of .33 (called *hard* or *low-TP words*), as the syllables they entail were also presented in other “words” embedded in the stream, to mimic closely what occurs in natural languages as the same syllable (e.g., “cur”), can appear in different words (e.g., “cur.va.ture”, “in.cur.sion” “re.oc.cur”). Moreover, the same participants performed the task, first under implicit conditions (i.e., without any instructions about the nonsense words embedded in the stream as typically observed in SL studies) and, afterwards, under explicit conditions (i.e., with the previous knowledge of the nonsense words embedded in the stream) to further analyze the role that the explicit knowledge play on SL. Electrophysiological (EEG) data were collected during the familiarization phases of each of the SL tasks, which allowed the authors to study the processes underlying SL, and not only the results of SL as typically obtained through the use of the 2-AFC task

performed after the familiarization phase (see [18,37], see also [45] for a discussion on the limitations of the 2-AFC task to test SL).

The results obtained with adult participants ([23,39,45]) showed that although the overall 2-AFC performance exceeded the chance level in both SL tasks (around 60%) it was nevertheless substantially below the levels observed in other SL studies¹. For example, accuracy rates averaged 74% in [46], 75% in [19] and 89% in [18]. The use of a higher number of “words” than used in previous SL studies (i.e., eight vs. four/six - see [18,19,46]) that were repeated fewer times (60 instead of 100 or more), along with the use of different types of “words” (high- and low-TP words) in the same stream, were pointed out by the authors [23] as potential explanations for the results. It is also important to note that although in that study the differences between high- and low-TP words failed to reach statistical significance both in the SL tasks performed under implicit and explicit conditions, in subsequent studies the authors found reliable signs of SL in the 2-AFC task performed under implicit conditions, but, surprisingly, only for the low-TP words. To account for this unexpected result, Soares et al. [45] called attention to an inevitable consequence that the manipulation of the TPs in their studies entailed. Indeed, because high-TP words were made of unique syllables that occurred only in specific “words”, conversely to low-TP words, whose syllables appeared in different “words”, this made the syllables of the high-TP words to occur three times less frequently than the syllables of words entailing the same syllables at different syllable positions (first, second and third) – i.e., the low-TP words. Thus, even though high- and low-TP words appeared exactly the same number of times during exposure to account for word frequency effects in processing, the fact that low-TP words entail syllables that occurred more often might have led participants when asked to decide which stimuli in a pair “sounded more familiar” (i.e., when asked to perform the 2-AFC task) to choose the ones that contained syllables that had occurred more frequently and that certainly generated higher levels of familiarity. The shift from a TP-based strategy to a syllable-frequency-based strategy these results seem to suggest is also in line with recent studies showing that the level of entropy interferes with the ease with which TPs/word-like units were extracted from the speech streams used in SL research (e.g., [36,51,52]).

Entropy is a term originally derived from thermodynamics to design the level of uncertainty in an information source ([53]) with less deterministic patterns showing higher levels of levels of entropy). The concept was recently applied to SL research by Siegelman et al. ([52]) as an index of stream learnability operationalized as the predictability of a given element (e.g., syllable) based on the probability distribution of all the elements in the stream based on the Markov’s formula – see also [36] for another implementation of entropy in SL research using Shannon’s formula. Entropy is thus a global statistic that computes distributional information of the input as a whole and that, together with TPs (a local statistic), has been assumed to affect SL processes and their results (see [29]). Previous studies suggest that the lower the level of entropy of a stream, the higher the learning rates (e.g., [36,51,52,54]). Thus, it is possible that the higher level of entropy of the streams used by Soares et al. ([23,39,40,45]) due to both the use of a high number of “words” than used in previous works and/or the use of “words” presenting different levels of predictability, might made the stream more complex to process and “words” harder to extract, hence justifying the lower rates of SL observed. Yet, it is also important to point out that there are also other studies suggesting that moderate levels of entropy facilitate SL (see [55–60]) as it provides the “right” level of variability to support pattern detection in line with the Goldilocks principle – SL is optimized in streams that are neither too predictable nor too unpredictable.

¹ The studies conducted with children ([39–42]) showed that performance was not significantly different from chance, which was in line with other studies indicating that the 2-AFC task is not suited for children below six years of age (e.g., [47,48,49] - see also [50] for similar findings in the context of the artificial grammar learning [AGL] task).

Still, the EEG data collected during the familiarization phase of the triplet-embedded paradigm by Soares et al. ([23,39,45]) with adult participants showed modulations in the N100 and, particularly, in the N400 ERP components considered to be signatures of SL in the brain (e.g., [18,21,22,46]). Specifically, the results revealed enhancements in the N100 component, suggesting an increased sensitivity to the regularities embedded in the input as exposure to the speech stream unfolded, and the recruitment of predictive mechanisms in the processing of speech streams (e.g., [61]). Importantly, in the N400 component, neural data showed larger amplitudes for the high- than for the low-TP “words” regardless of the conditions (implicit vs. explicit) under which the SL task was performed, suggesting facilitated access to high-TP representations in memory and/or more successful integration of high-TP representations into higher-order language structures, as expected to observe at the behavioral level. The inconsistency between the behavioral and EEG results might indicate, as we have been claiming, that the EEG data obtained from the familiarization phase and the behavioral data obtained from the test phase through the use of the 2-AFC task might tap into different processes and mechanisms. Indeed, besides being a post-learning task assessing SL only after exposure, it is also important to highlight that in the 2-AFC task participants are asked to make explicit judgments about regularities that are expected to be acquired implicitly (i.e., without intention and awareness), which might create not only a mismatch between the “mode of learning” and the “mode of assessing”, but also to allow other meta-cognitive factors to affect the results (for a discussion see [18,23] and [62]). Nonetheless, it is also important to consider that the disparity in the results observed by Soares et al. ([23,39,45]) could also stem from the complexity (entropy) of the streams used, even though understanding which levels of entropy enhance SL processes without overwhelming the cognitive system, as well as the factors that may influence them (e.g., prior knowledge), are crucial questions that remain largely overlooked in SL research.

This work was designed to shed light on this issue, i.e., on how the composition of the streams presented to participants both under implicit (without the previous knowledge of the “words” embedded in the stream) and explicit (with the previous knowledge of the “words” embedded in the stream) learning conditions could affect SL processes and their results. To that purpose, three types of auditory streams were constructed. Each stream contained four three-syllable nonsense words drawn from Soares et al.’s ([23]) study. Two of those streams were homogenous (*unmixed streams*), in the sense that they contained only one type of “words”, either high-TP (1.0) or low-TP (.33) words. The other type was heterogeneous (*mixed streams*), in the sense that half of the three-syllable nonsense words were high-TP (1.0) and the other half were low-TP words (.33), as in Soares et al. previous works ([23,39,40,45]) but using a lower number of three-syllable nonsense words (four) to approach closely previous studies (e.g., [1,18,19,46]) to which these results could be directly compared. The rationale behind this proposal was that if the number of words affected SL, as we expected, then we should observe higher levels of 2-AFC performance here than in the previous Soares et al. ([23,39,45]) works. Once homogeneous streams with low-TP words present higher levels of entropy (.48, according to Markov’s formula) than streams with high-TP words (.16, according to Markov’s formula)², 2-AFC performance was also expected to be better for the unmixed high-TP than for the unmixed low-TP stream condition. Importantly, if moderate levels of entropy facilitate SL as claimed

² These values were obtained through the computation of the Markov’s entropy formula, $i=1n p(i)j=1n p(i) \log p(j|i)$ taking TPs for all possible transitions (syllable pairs) in each of the streams into account. This formula sums, for each transition (i, j), the product of the distributional probability of i [$p(i)$], with the conditional probability of $i|j$ [$p(j|i)$], and its logarithm [$\log p(j|i)$]. Both a higher number of triplets and lower TPs increase entropy. A higher number of triplets increases entropy through the decrease of predictability in “word” boundaries since each “word” can be followed by $n-1$ “words”, or $n-2$ for low-TP “words”. Lower TPs increase the level of “surprisal” for the syllable transition. This concept proves useful for better detailing the predictability of a set of stimuli in a stream, giving us a more complete and unified assessment of predictability.

by the Goldilocks principle, performance should be better in the mixed condition (0.26, according to Markov's formula) than both in the unmixed low-TP and unmixed high-TP conditions, particularly in the SL tasks performed under implicit conditions, once participants cannot rely on prior knowledge to support SL.

Furthermore, if the presence of high- and low-TP words within the same stream encourages the use of a syllable frequency-based strategy rather than a syllable TP-based strategy, as suggested by Soares et al. ([39,45]), recognition rates would be expected to be higher for low-TP words than for high-TP words in the mixed stream condition. However, it is also plausible that with less complex streams, behavioral and neural results may converge, leading to better 2-AFC performance for high-TP than low-TP words, consistent with the neural-level findings reported by Soares et al. ([23,39,45]). In the same vein, if the streams' composition affected SL processes, modulations in the N100 and N400 components were expected to be observed across conditions. Specifically, enhancements in the N100 and N400 components, taken as the neural signatures of SL in the brain, should be observed in the mixed high-TP condition relative to the other stream conditions. Moreover, in the mixed condition, it is also possible that high-TP words produce this kind of effect, particularly under explicit conditions as observed in previous works.

2. Materials and Methods

2.1. Participants

Fifty-six undergraduate students ($M_{\text{age}} = 21.4$; $SD_{\text{age}} = 4.0$; 47 women) from the University of Minho participated in the study. The sample size was estimated based on previous studies. A similar sample size was also obtained from power analysis conducted using G*Power ([63]) to achieve 95% power for detecting a medium effect at a significance criterion of $\alpha = .05$ ($1 - \beta = 0.95$; $\alpha = 0.05$) for an effect size of (U) = 0.40 ($\eta^2 \sim 0.082$). All participants were native speakers of European Portuguese, with normal hearing and no reported history of learning or language disabilities and/or neurological problems. All were right-handed, as assessed by the Portuguese adaptation of the Edinburgh Handedness Inventory ([64,65]). Written informed consent was obtained from each participant. The local Ethics Committee approved the study.

2.2. Stimuli

The nonsense words were taken from Soares et al.'s ([23]). They were made of 32 unique Portuguese syllables recorded by a native speaker of European Portuguese with a duration of 300 ms each. The syllables were evenly distributed across two syllabaries (syllabary A and B – see Table 1) used either in the implicit or explicit versions of the SL tasks (counterbalanced across participants) to avoid interference. In each syllabary, the 16 syllables were concatenated into triplets presenting either TPs of 1.00 (high-TP “words”) or .33 (low-TP “words”). The high-TP “words” entailed syllables that only occurred in each of those words, such as “tucida” from syllabary A and “todidu” from syllabary B; while low-TP words entailed syllables that occurred in several words at different (initial, medial, and final) syllable positions, such the syllable “do” in the nonsense words “dotage” “tidomi” and “migedo” from syllabary A and the syllable “pi” in the nonsense words “pitegu” “tepime” “megupi” from syllabary B (see Table 1). The high-TP “words” were used to construct the unmixed high-TP streams, whereas the low-TP “words” were used to construct the unmixed low-TP streams. Two mixed streams (mixed A and mixed B) made of two of the four “words” of unmixed high-TP condition and two of the four “words” of the unmixed low-TP condition were also constructed for control. See Table 1 for an illustration of the type of streams constructed.

Table 1. Examples of the “words” and foils used per stream condition.

Type of stream	Syllabary A		Syllabary B	
	“Words”	Foils	“Words”	Foils
unmixed high-TP	tucida	tubago	todidu	tomabe
	bupepo	bucica	cegita	cedico
	modego	mopeda	gapabe	gagidu
	bibaca	bidepo	bomaco	bopata
unmixed low-TP	dotige	dogeti	pitegu	pigute
	tidomi	timido	tepime	temepi
	migedo	midoge	megupi	mepigu
	gemit	getimi	gumete	guteme
mixed A	migedo	gebado	megupi	gumapi
	gemit	mogeti	gumete	gagute
	modego	bimigo	gapabe	bornebe
	bibaca	mideca	bomaco	mepaco
mixed B	dotige	docimi	pitegu	tegiu
	tidomi	budoge	tepime	toteme
	tucida	tutipo	todidu	cepidu
	bupepo	tipeda	cegita	pidita

The nonsense words were concatenated with the Audacity® software (1999-2019) with a 50 ms interval between syllables as in previous studies ([39–42,45]). They were presented in a pseudo-randomized order, such that the same “word” or the same syllable would never appear consecutively in a row (i.e., situations such as “tucidatucida” or “tidomimigedo” were not allowed to occur) to avoid confounds (e.g., [1,18–20] see also [60]). In each stream, the “words” were presented 60 times in six blocks of 10 repetitions each lasting 4.2 min (around half a second per block). TPs across “word” boundaries were, therefore, .33 for the unmixed high-TP and mixed conditions, and .50 for the unmixed low-TP condition. Each stream was also edited to include a superimposed chirp sound (a 0.1 s sawtooth wave sound from 450 to 1,450 Hz) in 10% of the syllables to provide participants with a cover task (i.e., a chirp detection task) during exposure, as in previous studies (e.g., [23,39,40,45]). The chirp was included in all “words” counterbalanced across syllable positions to prevent confounds (see Figure 1 ahead).

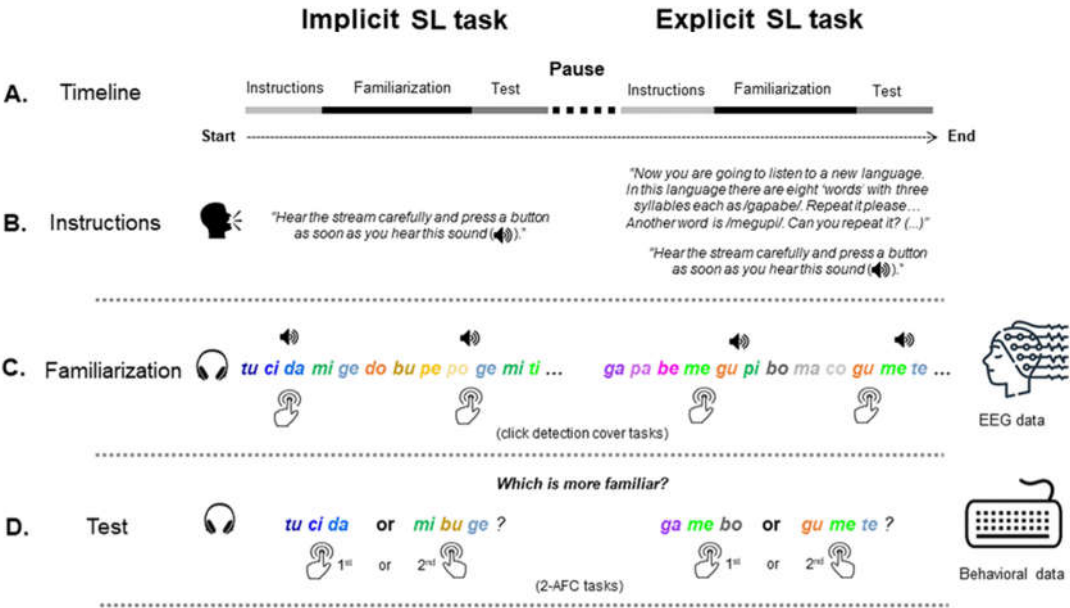


Figure 1. Experimental setup. Panel A illustrates the timeline of the experimental procedure in which the implicit and, subsequently, the explicit SL tasks were performed. Each SL task comprised three parts: instructions (Panel B), familiarization phase (Panel C), and test phase (Panel D). Each task was initiated with specific instructions that determined the conditions under which the SL task was performed: implicit instructions (i.e., without knowledge of the stimuli or the structure of the stream) and explicit instructions (i.e., with explicit knowledge or pre-training on the ‘words’ presented in the stream). In the familiarization phase of both tasks (Panel C) during which EEG data were collected, participants were presented with a continuous auditory stream made of four three-syllable nonsense words of either two high-TP “words” and two low-TP “words” (mixed condition) as depicted on the Figure, or four high-TP “words” (unmixed high-TP conditions), or four low-TP “words” (unmixed low-TP conditions) or with chirp sounds (speaker icon on the Figure) superimposed over specific syllables. The chirp sounds could emerge at any of the three-syllable positions of the “words”, which precluded its use as a cue for stream segmentation. During this phase, participants had to perform a chirp detection task. Then, a test phase (Panel D) consisting of a 2-AFC task asked participants to indicate which of the three-syllable sequences (a “word” and a foil) sounded more familiar, considering the stream heard during the familiarization phase.

For the 2-AFC tasks, 16 foils were created for each type of stream. The foils were created from the same set of syllables that made the “words” in each stream and syllabary (see Table 1). The syllables in the foils were used with the same frequency and syllable positions as the syllables in the “words”, though presented in an order not presented before. For example, the syllable “do” which appeared three times at different syllable positions (initial, medial, final) in three different “words” from the unmixed low-TP condition, also appeared three times at these syllable positions in the foils (e.g., “dogeti”, “midoge”, “timido” – see Table 1). Note, however, that due to stimuli restrictions (i.e., to the number of syllables used to make the four high-TP and the four low-TP “words” in each syllabary – four vs. 12), the foils in the unmixed high-TP and the mixed conditions present TPs of 0 as the order of the syllables they entailed was entirely new. However, in the unmixed low-TP condition, the foils present TPs of .25, once the number of syllables available and the need to not repeat the same “word” consecutively did not preclude the possibility of a given syllable pair presented in the foils to have occurred during exposure across “word” boundaries (e.g., the syllable pair “geti” presented in the foil “dogeti” could have also appeared when the nonsense word “dotige” was followed by the nonsense “word” “tidomi” in the familiarization phase).

Four lists of materials were created for each type of stream to counterbalance syllables across positions in each syllabary (eight lists per type of stream, 32 lists in total). Participants were randomly assigned to one list of one of the three types of streams (unmixed high-TP, unmixed low-TP, or mixed A or B). Sixteen participants were assigned to the unmixed high-TP condition, 16 to the unmixed low-TP condition, and 24 to the mixed condition (12 participants to the A and 12 to the B version). Note that when assigned to a given stream condition, participants made the implicit and the explicit versions of the SL task under the same condition (either unmixed high-TP, unmixed low-TP, or mixed), though using stimuli from syllabary A or syllabary B to avoid carry-over effects, as mentioned before. The high number of participants assigned to the mixed condition was due to the need to increase the statistical power of the analysis to be conducted, as we intend to explore the effects of the type of “word” in the results (note that in this condition we only use two items per type of “word”).

2.3. Procedure

Participants were first presented with the implicit version of the SL task and, then, with the explicit version of an analogous SL task (see **Figure 1**). In the implicit version, participants were instructed to pay attention to the auditory stream (a sequence of syllables) presented at 60 dB SPL via binaural headphones, because occasionally a deviant sound (i.e., a click) would appear, and their task would be to detect it as soon and accurately as possible by pressing the spacebar from the computer

keyboard. This functioned as a cover task to ascertain participants paid adequate attention to the stream. Following exposure, participants were asked to decide as accurately as possible which of two auditory stimuli (one “word” and one foil) “sounded more like” the stimuli presented before (i.e., to perform a 2-AFC task). The 2-AFC comprised 16 trials in which each of the four “words” presented during familiarization was paired with the four foils. As in Soares et al. ([39]), this option sought to minimize the number of times each “word” and foil was presented once Soares et al. ([45]) have shown that increasing the number of 2-AFC trials by repeating the same stimuli (“words” and foils) several times throughout the task has detrimental effects on SL performance once foils tend to emerge as perceptual units interfering with a correct “word” discrimination (see ([45] for details).

In the 2-AFC task, each trial began with the presentation of a fixation point (cross) for 1,000 ms after which the first stimulus (“word” or foil) was presented, followed by the second stimulus. A 500 ms inter-stimulus interval separated the presentation of the stimuli. The next trial began as soon as participants made a response or 10 sec had elapsed. The 16 trials were presented in four blocks of four trials each. In each block, the order (first or second) by which the stimuli were presented was controlled, so that in half of the trials, half of the “words” were presented first and in the other half the reverse (counterbalanced across blocks). The trials in each block, as well as the blocks, were randomly presented to the participants.

After a brief interval, participants underwent the explicit version of the SL task. This version followed the same procedure except that, before the new familiarization phase began, the stimuli (i.e., the nonsense words) that they would listen to were taught. Specifically, participants were told that they would be listening to some “new words” from another foreign language. Then, each of the four new “words” was presented auditorily (one by one) and participants were asked to repeat each of them correctly. As in the implicit version of the task, during familiarization, participants were asked to press a button (i.e., the spacebar of the computer keyboard) whenever they heard the click sound to ensure adequate attention to the stream. After familiarization, participants performed another 2-AFC task that mimicked the one used in the implicit version of the SL task. The procedure took about 60 min to complete per participant. Figure 1 depicts a visual summary of the experimental design adopted in the mixed condition, as an example. In all the other conditions the procedure was exactly the same except for the use of other types of streams (unmixed high-TP and unmixed low-TP streams).

2.4. EEG Data Acquisition and Processing

Data collection was performed in an electric-shielded, sound-attenuated room at the facilities of the Psychological Neuroscience Lab (School of Psychology, University of Minho). Participants were seated in a comfortable chair, one meter away from a computer screen. During the familiarization phase, EEG data was recorded with a 64 channels BioSemi Active-Two system (BioSemi, Amsterdam, The Netherlands) according to the international 10-20 system and digitized at a sampling rate of 512 Hz and downsampling to 256 Hz. Electrode impedances were kept below 20 k Ω . EEG was re-referenced offline to the algebraic average of mastoids. Data were filtered with a bandpass filter of 0.1 - 30 Hz (zero phase shift Butterworth) plus a notch filter of 50 Hz. Interpolation was performed for electrodes with noise or flat. Independent component analyses (ICA) were performed to remove stereotyped noise (mainly ocular movements and blinks) by subtracting the corresponding components. ERP epochs were time-locked to the nonsense words’ onset, from -300 ms to 1,200 ms (baseline correction from -300 to 0 ms). After that, epochs containing artefacts (i.e., with amplitudes exceeding ± 100 μ V) were excluded. EEG data processing was conducted with Brain Vision Analyzer, version 2.1.1. (Brain Products, Munich, Germany).

2.5. Data Analyses

Behavioral (2-AFC) and ERP data analyses were performed using the IBM-SPSS® software (Version 27.0). All participants showed a click detection performance above 90% during

familiarization, except one from the mixed stream condition, which was excluded from the analyses. For behavioral data, the % of correct responses was computed for each of the 2-AFC tasks (implicit and explicit) and separately per stream (unmixed high-TP, unmixed low-TP and mixed – mixed A and mixed B collapsed). One-sample *t*-tests against the chance level were conducted on the data of each of the SL tasks and streams to determine whether the performance was significantly different from chance (50%). A repeated measure analysis of variance (ANOVA) using Type of stream (unmixed high-TP, unmixed low-TP, or mixed) as a between-subject factor, and SL task (implicit vs. explicit) as a within-subject factor was also conducted to analyze if 2-AFC performance was significantly different across groups. For the mixed stream condition, another ANOVA was conducted to ascertain whether the 2-AFC performance was significantly different across the type of “words” (high- vs. low-TP) and the conditions under which the SL task was performed (implicit vs. explicit).

Individual ERPs were averaged separately per SL task (implicit and explicit), stream condition (unmixed high-TP, unmixed low-TP, or mixed – mixed A and mixed B collapsed), and length of exposure (*half 1*: block#1, block#2, block#3 vs. *half 2*: block#4, block#5, block#6). As in previous studies we chose to analyze neural data into two different halves to further examine if the brain responses to the different conditions would emerge as a function of the amount of exposure to the input ([23,39,40]). Due to artifact rejection (rejected more than 50% of the trials) six participants were excluded from the unmixed high-TP stream condition in the SL task performed under implicit conditions and four in the SL task performed under explicit conditions; four participants were excluded from the unmixed low-TP stream condition in the SL task performed under implicit conditions and one in the SL task performed under explicit conditions; and six participants were excluded from the unmixed high-TP stream condition in the SL task performed under implicit and explicit. After artefact rejection, the average accepted trials by condition and group was 83% (178 trials).

Based on previous SL studies (e.g., [18,21–23,39,40,46]) and on the general inspection of the data peak amplitudes were measured for the N100 (60–120 ms) and N400 (350–450 ms) components, taken as the neural signatures of words’ segmentation in the brain. Nonetheless, since a general inspection of the neural results suggested that the positivity observed within the 120–220 ms time window, corresponding to the P200 component, might reveal contrasts of interest, we decided to further examine this component, even though it was not initially considered. Previous SL studies have also reported an enhancement of this component in auditory SL tasks, which has been associated with an early perceptual learning of the regularities embedded in the speech streams and attention allocation [66–69]. To account for the topographical distribution of the abovementioned EEG modulations, peak amplitudes’ values were obtained for the frontocentral regions of interest (ROI; F1, Fz, F2, FC1, FCz, FC2, C1, Cz, and C2) where amplitudes were maximal for the N100, P200 and N400 components.

Repeated measure ANOVAs were conducted on the neural data obtained for the target components in two sets of analyses. In the first, considering the type of stream (unmixed high-TP, unmixed low-TP, or mixed) as a between-subject factor, and length of exposure (*half 1* vs. *half 2*) as within-subject factors. Note that unlike the behavioral analyses and EEG results reported in previous studies ([23,39,40]), here we opted to analyze the SL tasks performed under implicit and explicit conditions separately. This decision was driven by the high number of participants excluded from the EEG analysis—primarily in the unmixed high-TP condition due to artifact rejection. Introducing the SL task condition (implicit vs. explicit) as an additional within-subject factor in the ANOVA would have limited the analysis to only eight participants in the unmixed high-TP condition, significantly reducing statistical power. In a second set of ANOVA analyses, focused on the mixed stream condition, we included SL task (implicit vs. explicit), type of word (high- vs. low-TP), and length of exposure (*half 1* vs. *half 2*) as within-subject factors. This approach was feasible because, in this case, neural data from 17 participants were available for both implicit and explicit SL tasks – six due to artifact rejection, and one due to click detection performance below 90%.

Both for behavioral and ERP data, main and interaction effects that reached statistical or marginal significance levels ($p < .05$ or $p < .08$, respectively) in comparisons of interest are reported. The Greenhouse–Geisser correction for nonsphericity was used when appropriate. Post-hoc tests for multiple comparisons were adjusted with Bonferroni correction. In such cases, the p -values reported were the ones obtained after the Bonferroni corrections were automatically applied (i.e., the adjusted p -values) by the IBM-SPSS® software (Version 29.0). Measures of effect size (Eta squared, η^2) and observed power (pw) for a single effect are reported in combination with the main effects of the condition.

3. Results

3.1. Behavioral Data

The mean percentages of correct responses obtained for the 2-AFC tasks performed under implicit and explicit conditions per type of stream are presented in Table 2.

Table 2. Mean percentages of correct responses for 2-AFC tasks performed under implicit and explicit conditions per type of stream.

Type of stream	SL Task	
	Implicit	Explicit
Unmixed high-TP	61.3 (11.5)	71.5 (12.3)
Unmixed low-TP	55.9 (8.7)	64.1 (13.4)
Mixed	62.2 (15.1)	69.6 (17.4)

The results from the one-sample t -tests against chance level showed that participants revealed an above-chance 2-AFC performance in all the conditions, unmixed high-TP: implicit SL task, $t(15) = 3.96$, $p < .001$, explicit SL task, $t(15) = 7.00$, $p < .001$; unmixed low-TP: implicit SL task, $t(15) = 2.70$, $p = .008$, explicit SL task, $t(15) = 4.20$, $p < .001$; mixed: implicit SL task, $t(22) = 3.88$, $p < .001$; explicit SL task, $t(22) = 5.39$, $p < .001$. Moreover, in the mixed condition, the one-sample t -tests against chance level showed that performance exceeded the chance levels for both types of “words” in the SL tasks performed under implicit and explicit conditions (see **Table 3**), high-TP “words”: implicit SL task, $t(22) = 3.29$, $p = .002$, explicit SL task, $t(22) = 5.54$, $p < .001$; low-TP “words”: implicit SL task, $t(22) = 2.24$, $p = .018$, explicit SL task, $t(22) = 3.57$, $p < .001$.

Table 3. Mean percentages of correct responses for 2-AFC tasks performed under implicit and explicit conditions in the mixed stream condition per type of “word”.

Type of word	SL Task	
	Implicit	Explicit
High-TP	65.8 (23.0)	75.0 (21.7)
Low-TP	58.7 (18.6)	64.1 (19.0)

Additionally, the results from the repeated measures ANOVAs indicated that only the main effect of SL task reached a statistically significant level, $F(1, 52) = 12.54$, $p < .001$, $\eta^2 = .194$, $pw = .935$. This effect showed that performance was better when the 2-AFC tasks were performed under explicit than implicit learning conditions regardless of the type of stream considered (68.4 vs. 59.8, respectively).

Finally, the results of the ANOVA conducted on the data obtained from the mixed condition showed a main effect of word type, $F(1, 22) = 5.32$, $p = .031$, $\eta^2 = .195$, $pw = .597$, indicating that high-TP words were recognized more accurately than low-TP words (70.4 vs. 61.4, respectively). No other main or interaction effects reached statistical significance.

3.2. Electrophysiological Data

3.2.1. N100

In the first set of analyses, contrasting the three types of streams, the ANOVA revealed that the only effect that reached a statistically significant level was the effect of length of exposure, $F(1,38) = 5.234$, $p = .028$, $\eta_p^2 = .121$, $pw = .606$ in the SL task performed under implicit conditions. In the SL tasks performed under explicit conditions, the results failed to show any statistically significant main or interaction effect. **Figure 2** depicts the neural responses (mean amplitudes' values and topographical maps) observed in this time window (the grey-shadowed rectangle) for SL tasks performed under implicit (left panel) and explicit (right panel) learning conditions in the first (solid lines) and second (dotted lines) halves for each of the streams (yellow: unmixed high-TP condition; red: unmixed low-TP condition; and orange: mixed condition).

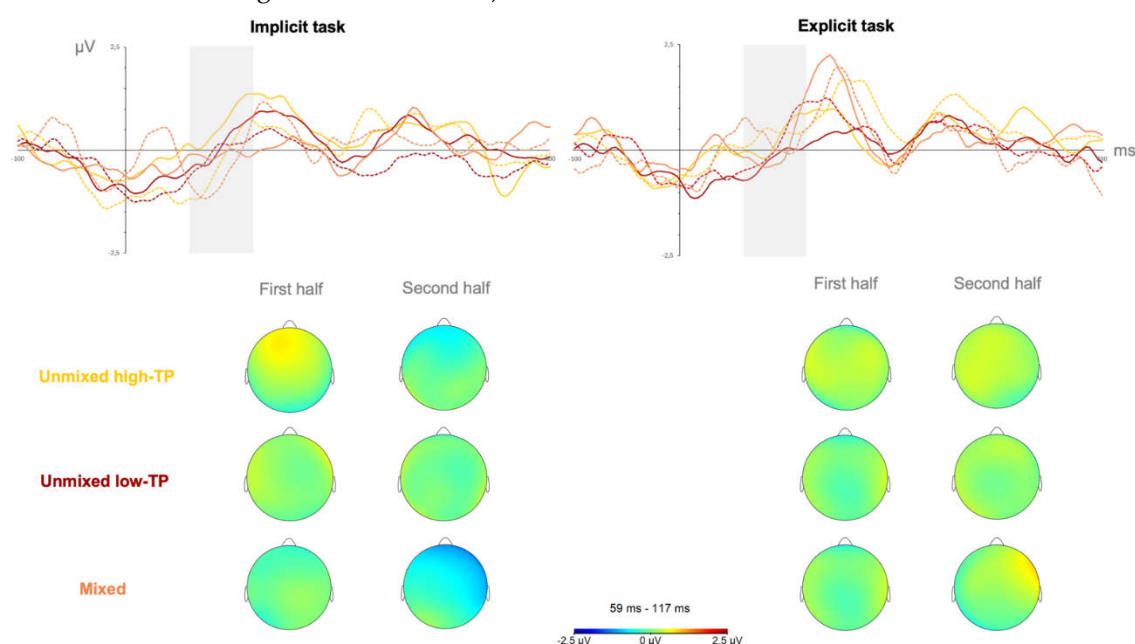


Figure 2. Neural responses in the first and second halves of the SL tasks performed under implicit (left panel) and explicit (right panel) conditions in the N100 component. Yellow solid line = unmixed high-TP stream, first half; yellow dotted line = unmixed high-TP stream, second half; red solid line = unmixed low-TP stream, first half; red dotted line = unmixed low-TP stream, second half; orange solid line = mixed stream, first half; orange dotted line = mixed stream, second half.

The length of exposure effect observed in the implicit SL tasks in this time window showed a larger amplitude in the second than in the first halves of the familiarization phase, as can be inferred from Figure 2 by contrasting, in the grey shadowed rectangle from the left panel, the solid lines with dotted lines in each of the stream conditions.

In the second analysis, focused on data from the mixed stream condition, the ANOVA showed that the main effect of word type reached a marginally statistically significant level, $F(1,16) = 4.111$, $p = .060$, $\eta_p^2 = .204$, $pw = .478$ in this time window. **Figure 3** depicts the neural responses (mean amplitudes' values and topographical maps) from the mixed stream condition for the high-TP (light blue lines) and the low-TP (dark blue lines) words in the SL tasks performed under implicit (solid lines) and explicit (dotted lines) conditions in the N100 (first grey shadowed rectangle) and P200 (second grey shadowed rectangle) components regardless of the length of exposure to the streams (first and second halves collapsed).

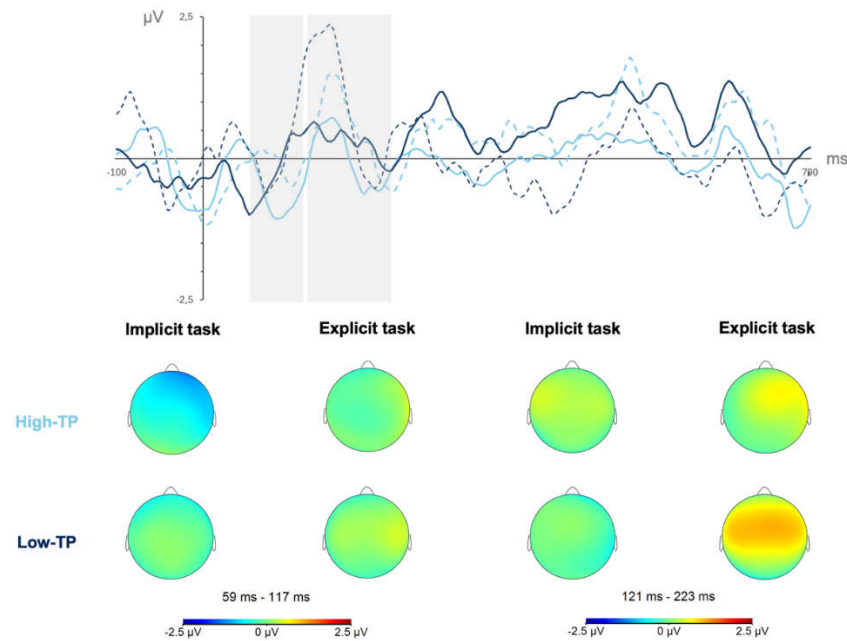


Figure 3. Neural responses from the mixed stream condition in the N100 (first grey shadowed rectangle) and P200 (second grey shadowed rectangle) components for the high-TP (light blue lines) and for the low-TP (dark blue lines) words in the SL tasks performed under implicit (solid lines) and explicit (dotted lines) conditions.

The type of word effect observed in the N100 time window, revealed a larger amplitude for the high- than for the low-TP words regardless of the conditions (implicit and explicit) under which participants performed the SL tasks, as can be inferred from Figure 3 by contrasting the solid and dotted light blue lines with the solid and dotted dark blue lines in the first grey shadowed rectangle.

3.2.2. P200

The results from the first ANOVA, comparing the three type of streams, revealed that the main effect of stream reached a statistically significant level, $F(2,42) = 3.842$, $p = .029$, $\eta_p^2 = .155$, $pw = .665$ in the SL tasks performed under explicit conditions. In the SL tasks performed under implicit conditions the results failed to show any statistically significant main or interaction effect. **Figure 4** depicts the neural responses (mean amplitudes' values and topographical maps) observed in the P200 (first grey-shadowed rectangle) and N400 (second grey-shadowed rectangle) components for the SL tasks performed under implicit (solid lines) and explicit (dotted lines) conditions per stream (yellow: unmixed high-TP condition; red: unmixed low-TP condition; and orange: mixed condition) regardless of the length of exposure (first and second halves collapsed).

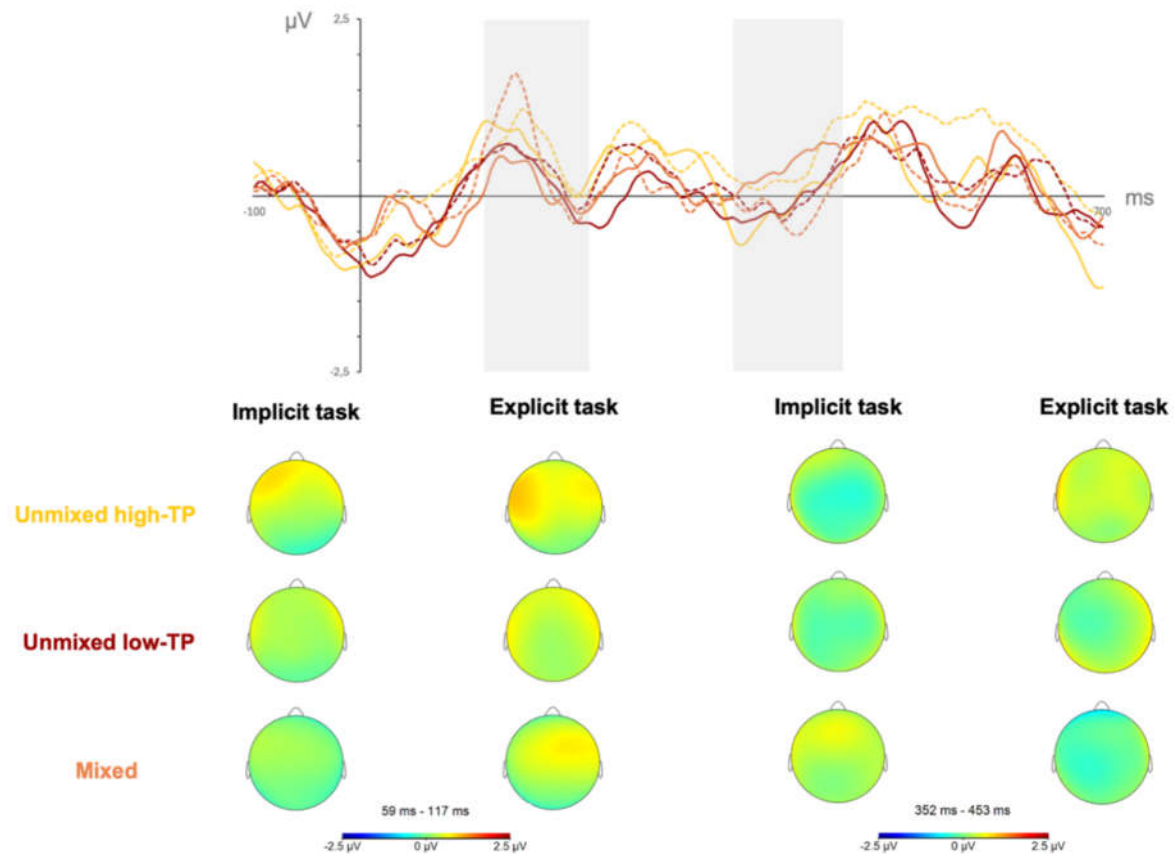


Figure 4. Neural responses in the implicit and explicit SL tasks in the P200 (first grey-shadowed rectangle) and N400 (second grey-shadowed rectangle) components for the SL tasks performed under implicit (solid lines) and explicit (dotted lines) per stream condition: Yellow solid line = Unmixed high-TP stream, implicit; Yellow dotted line = Unmixed high-TP stream, explicit; Red solid line = Unmixed low-TP stream, implicit; Red dotted line = Unmixed low-TP stream, explicit; Orange solid line = Mixed stream, implicit; Orange dotted line = Mixed stream, explicit.

The stream effect observed in the P200 component showed a larger amplitude in the mixed than in the unmixed low-TP condition ($p = .031$), irrespective of whether participants were exposed to the auditory streams under implicit or explicit conditions, as can be inferred by contrasting, the orange solid and dotted lines (mixed condition) with the red solid and dotted lines (unmixed low-TP condition), in the first grey shadowed rectangle of **Figure 4**.

In the second analysis, restricted to the neural data from the mixed stream condition, the ANOVA revealed that the type of SL task main effect reached a marginally statistically significant level. The task effect observed in this time window revealed a larger amplitude for the SL tasks performed under explicit than under implicit conditions, as can be inferred by contrasting in the second grey shadowed rectangle of **Figure 3** the light and dark blue dotted lines (explicit conditions) with the dark blue solid lines (implicit conditions).

3.2.3. N400

The results from the first ANOVA comparing the three type of streams revealed a significant main effect of type of stream, $F(2,42) = 4.387$, $p = .019$, $\eta_p^2 = .273$, $pw = .727$ in the SL tasks performed under explicit conditions, similar to the findings in the P200 time window. However, in this case, the *post-hoc* comparisons revealed that the difference across stream conditions reached statistical significance when comparing the mixed with the unmixed high-TP stream conditions, with the first exhibiting a larger amplitude than the former ($p = .017$), as can be inferred by contrasting the orange solid and dotted lines (mixed condition) with the yellow solid and dotted lines (unmixed high-TP

condition), in the second grey shadowed rectangle of **Figure 4**. The results of the first ANOVA conducted on neural data from the implicit learning conditions, as well as from the second ANOVA conducted on the neural data from the mixed stream condition, did not reveal any statistically significant main or interaction effects.

4. Discussion

The present study examined how the composition of the speech streams affected SL processes and their results. We also aimed to ascertain how the prior knowledge of the three-syllable nonsense words embedded in the speech streams modulated these results. To that purpose, three types of speech streams containing either four high-TP words (i.e., three-syllable nonsense words with TP = 1; unmixed high-TP condition), four low-TP words (i.e., three-syllable nonsense words with TP = .33; unmixed low-TP condition) or two high-TP and two low-TP words (mixed condition) were presented to participants, first, under incidental (implicit), and, subsequently, under intentional (explicit) conditions as in Soares et al. ([23,39–42]) previous SL studies. Neural and behavioral data were collected during the familiarization and test phases of each of the SL tasks, respectively. This paper is, to the best of our knowledge, the first directly examining these issues with important methodological and theoretical implications. The findings can contribute not only to deepening our understanding of how SL mechanisms work in a wide range of situations –approaching closely what might occur in natural languages acquisition; but also to shed light on the type of statistics (local statistics vs. global statistics) participants are sensitive and use to extract word-like units from the continuous auditory streams.

The results obtained were straightforward and can be summarized as follows: (i) behavioral (2-AFC) signs of SL were observed for the three types of streams and learning conditions, particularly in the SL task performed under explicit conditions, as anticipated; (ii) although 2-AFC performance was numerically higher (and very similar) in the mixed and unmixed high-TP conditions than in the unmixed low-TP condition both in the SL tasks performed under implicit and explicit conditions, the 2-AFC differences across stream conditions failed to reach statistical significance; (iii) still, the 2-AFC data obtained from the mixed stream condition revealed a word type effect indicating better recognition rates for the high- than for the low-TP words; (iv) even though the behavior data failed to show any statistical significant stream effect, the neural data showed critically modulations both in the N400 component, foreseen from the outset, and in the additional P200 component showing a larger amplitude in the mixed vs the unmixed high-TP condition in the former and a larger amplitude in the mixed vs the unmixed low-TP condition in the latter, suggesting, in both cases, facilitative processing of the mixed over the other stream conditions; (v) additionally, the neural data revealed a length of exposure effect in the N100 component, indexed by a larger amplitude in the second half than in the first half of the SL tasks performed under implicit conditions, regardless of the type of stream, suggesting, importantly, that participants were able to extract the regularities embedded in the speech streams as exposure unfolded in all the stream conditions as expected; (vi) marginally significant word type effect was also observed in the N100 component for the mixed stream condition, with high-TP words eliciting larger amplitudes than low-TP words; (vii) although direct comparisons of neural responses between SL tasks were limited to the mixed condition (due to the high number of participants excluded from the EEG analyses because of artifact rejection, particularly in the unmixed high-TP condition), the results from the separate analyses for SL tasks performed under implicit and explicit conditions revealed that the stream effect observed in the P200 and N400 components emerged only under explicit learning conditions, hence suggesting an indirect SL task effect in the expected direction; and (viii) that a significant (albeit marginally) SL task effect was observed in the mixed stream condition as indexed by a larger P200 amplitude in the SL task performed under explicit than in implicit learning conditions, as expected.

These results are interesting and globally provide support to our hypotheses. Indeed, when the number of words embedded in the streams is halved (from eight in Soares et al. ([23,39–42]) previous

works to four) an increase in 2-AFC performance was observed, particularly in the SL tasks performed under implicit conditions, as expected. Specifically, a comparison between the results obtained here and those reported by Soares et al. ([39])— these works are more directly comparable since both used 16 trials in the 2-AFC post-learning task - revealed notable differences. When considering overall 2-AFC performance (i.e., regardless of word type), participants improved their performance by 9.5% in the implicit SL tasks (52.8% in Soares et al. [39] vs 62.2% here) and by 6.1% in the explicit SL tasks (63.5% in Soares et al. [39] vs 69.6% here). It is also worth noting that when we consider the type of word in the analyses, it is readily apparent that the increase in the 2-AFC performance was more pronounced for the high-TP than for the low-TP words. Indeed, the comparison of the results obtained here in the mixed stream condition with those obtained by Soares et al.'s [39] revealed that the 2-AFC performance for the high-TP words increased by 18.7% in the implicit SL tasks (47.1% in Soares et al.'s [39] work vs 65.8% obtained here), and by 13.6% in the explicit SL tasks (61.4% in Soares et al.'s [39] vs 75% obtained here). For the low-TP words, however, the comparisons revealed a quite similar pattern of results across studies with a difference of 0.3% in the SL tasks performed under implicit conditions (58.4% in Soares et al.'s [39] work vs 58.7% obtained here), and a difference of 1.4% in the SL tasks performed under explicit conditions (65.5% in Soares et al.'s [39] work vs. 64.1% obtained here).

In the case of the unmixed streams, the comparisons of the results obtained by in Soares et al.'s [39] for the high-TP words and those obtained here for the unmixed high-TP stream condition revealed an increase by 14.2% in the SL tasks performed under implicit conditions (47.1% in Soares et al.'s [39] work vs. 61.3% obtained here), and an increase by 10.1% in the SL tasks performed under explicit conditions (61.4% in Soares et al.'s [39] work vs. 71.5% obtained here). In the case of unmixed low-TP streams, however, the comparisons revealed a decrease of 2.5% in the SL tasks performed under implicit conditions (58.4% in Soares et al.'s [39] work vs. 55.9% obtained here) and a decrease of 1.4% in the SL tasks performed under explicit conditions (65.5% Soares et al.'s [39] work vs. 64.1% obtained here). Taken together, these findings indicate that the complexity (entropy) of the auditory streams presented to participants in the triplet-embedded paradigm strongly impacts the recognition of the high-TP words both when the SL tasks were performed under implicit and explicit conditions, which also agrees with the view that the representations generated under incidental and intentional learning conditions are not immune to interference as several authors have been claiming (e.g., [45,70–73]). Nevertheless, it is also important to note that the relatively low 2-AFC performance observed in the unmixed low-TP condition relative both to the mixed and the unmixed high-TP conditions could arise not only from the fact that it presents the higher level of entropy (.48), possibly overwhelming the cognitive system, but also from the fact that the foils used in the 2-AFC task present higher TPs (.25) than the foils used in the unmixed high-TP and stream mixed conditions (TP = 0) - due to the reduced number of syllables used in that condition, as mentioned before (see the Stimuli section); which may have made the foils more difficult to discriminate. Thus, the stream conditions more directly comparable, at least at a behavioral level of analysis, are the unmixed high-TP and the mixed conditions, which failed to show any statistically significant difference even though the 2-AFC performance was 6.3% higher in the mixed condition over the unmixed low-TP condition and 0.9% higher in the mixed condition over the unmixed high-TP condition.

Crucially, the results obtained from the EEG data collected during the exposure phase to the speech streams reached statistical significance. Indeed, the neural results revealed modulations as a function of the stream type not only in the N400 component, as anticipated from the outset, but also in the P200 component, which was explored additionally. In both cases, the findings suggest facilitative processing of the mixed stream condition over the unmixed low-TP stream condition (in the case of the P200 component) and of the mixed stream condition over the unmixed high-TP stream condition (in the case of the N400 component). These findings support the view that moderate levels of entropy enhance the detection of regularities embedded in the input and facilitate the extraction of word-like units from speech streams, extending previous evidence that has predominantly been demonstrated behaviorally (see [55–60]). By balancing predictability and variability, the mixed

streams used in our study appear to have created the optimal conditions to engage attentional mechanisms without overwhelming the system for SL as indexed by these results. These findings also support the view that results observed in lab SL research can be generalized to natural language acquisition contexts - which are characterized by higher levels of entropy than those typically used in lab settings; and that SL seems to be, indeed, a powerful mechanism that enables individuals to crack the language code and acquire language so quickly and effortlessly (see [2–10,15] see also [23,39–42]). Notably, these results were observed even if the neural data indicated that participants were sensitive to the regularities embedded in all the speech streams - as reflected by the main effect of length of exposure observed in the N100 component; and if behavioral signs of SL were observed across all stream conditions. The absence of behavioral differences across stream conditions is not surprising and aligns with several other studies claiming that the 2-AFC task is not suitable for testing SL, as it presents low sensitivity to detect subtle differences as typically used in SL experiments (see [15,45,49]). Additionally, 2-AFC differences across stream conditions may not be immediately apparent due to other factors such as task complexity or the time course of learning. These findings highlight the importance of using alternative measures to assess SL, particularly those sensitive to the time course of processing, such as EEG, during exposure to the speech streams, as used in an increasing number of SL studies ([18–20,22,23,33,35,39–42,46,47,51,68]).

Nevertheless, the 2-AFC results observed in the mixed condition, revealed that participants were significantly better at recognizing high-TP than low-TP words both under implicit and explicit conditions. These findings are not in accordance with the results observed previously by Soares et al. ([23,39]), who did not find reliable signs of SL for the high-TP words in the implicit learning condition, as mentioned. The fact that high-TP words had better recognition rates than low-TP words are, however, in line with the neural data observed by Soares et al. ([23,39]), who found greater N400 amplitudes for high than for the low-TP words, suggesting facilitated access to representations in memory and/or more successful integration of those representations in higher-order language structures. It is thus possible that the lower complexity of the streams used in the current work had allowed a more efficient extraction of the statistical regularities embedded in the input and the formation of more stable representations of the perceptual units (“words”) in long-term memory, hence avoiding the shift from a TP-based strategy to a syllable-frequency-based strategy that the complexity of the streams used by Soares et al. ([39,45]) seems to have stimulated.

The behavioral advantage of the high-TP words over the low-TP words observed in the mixed stream condition also agrees with the marginally statistically significant word type effect observed for this stream condition in the N100 component, suggesting that words with higher TPs are more easily detected and processed than words with lower TPs, potentially serving as “anchors” to facilitate the detection of the other words (i.e., the lower TP words) embedded in the stream. This finding replicates the results observed by Soares et al. ([39]) with adult participants, even though in an earlier time window, possibly due to the lower complexity of the streams used here, which included half of the words used by Soares et al. ([39]). It is also worth noting that this effect was observed when subjects were provided with prior knowledge about the regularities (words) embedded in the input, suggesting that explicit learning mechanisms are already at play at this early stage of processing, easing the decoding of the statistical structure of the sequence of syllables presented during the familiarization phase, particularly for those presenting high levels of predictability. The boost that previous knowledge seems to have on the extraction of regularities embedded in the mixed stream condition is also observed in the N400 component, taken as an index of the emergence of a pre-lexical trace of “words” in the brain (e.g., [21–23,39–42]). Specifically, the marginally statistically significant effect of SL task observed for the mixed stream condition in this time window revealed that words presented under explicit learning conditions elicited larger amplitudes than words presented under implicit conditions regardless of the type of word. This result, also observed in other studies (e.g., [19,20,23,39–41]), suggests a more successful segmentation of the speech streams into perceptual units (words) when “extra” (metalinguistic) information about

the to-be-learned regularities were provided to participants, which also agrees with the behavioral data obtained.

In conclusion, while the results obtained here are interesting and contribute to existing research by demonstrating that speech streams with moderate entropy levels create better conditions for SL—likely by striking a balance between predictability and variability to facilitate the extraction of regularities embedded in the input—many questions remain unanswered. For instance, it is still unclear what the upper limit of entropy is before the system becomes overloaded and how the level of entropy tolerated by the system may vary depending on factors such as the length of exposure to the stream or the knowledge participants acquire during this exposure, even if implicitly. Future research should, therefore, address these open issues, which will contribute to getting a deeper and more comprehensive understanding of how SL mechanisms operate under a wide range of situations. This would definitively help to bridge the gap between results obtained in controlled laboratory settings and the dynamics of “real-world” learning contexts, ultimately shedding light on the processes through which we acquire language so rapidly and effectively despite its inherent complexity.

Author Contributions: Conceptualization, Ana Paula Soares, Helena Mendes Oliveira and Dario Paiva; methodology, Ana Paula Soares, Helena Mendes Oliveira and Dario Paiva; software, Dario Paiva and Alberto Lema; formal analysis, Ana Paula Soares, Helena Mendes Oliveira, Alberto Lema, Diana R. Pereira, and Ana Cláudia Rodrigues; data curation, Helena Mendes Oliveira, Alberto Lema, Diana R. Pereira, and Ana Cláudia Rodrigues; writing—original draft preparation, Ana Paula Soares; writing—review and editing, Ana Paula Soares, Helena Mendes Oliveira, Alberto Lema, and Diana R. Pereira; visualization, Alberto Lema, Diana R. Pereira, and Ana Cláudia Rodrigues; project administration, Ana Paula Soares, and Helena Mendes Oliveira. All authors have read and agreed to the published version of the manuscript.

Funding: This work was conducted at the Psychology Research Centre (CIPsi/UM) School of Psychology, University of Minho, and by the Grant (FCT, 2022.05618.PTDC) supported by the Portuguese Foundation for Science and Technology (FCT).

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Ethics Committee for Research in Social and Human Sciences (CEICSH) from University of Minho with protocol code 096/2023 signed on 2023.08.01.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Data from this study can be downloaded at https://osf.io/g8xem/?view_only=250d9e5fc1344022b6818e98bdfd1055.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Saffran, J. R.; Aslin, R. N.; Newport, E. L. Statistical learning by 8-month-old infants. *Science* **1996**, *274*, 1926–1928. <https://doi.org/10.1126/science.274.5294.1926>
2. Breen, E.; Pomper, R.; & Saffran, J. Phonological learning influences label-object mapping in toddlers. *J. Speech Lang. Hear. Res.* **2019**, *62*(6), 1923–1932. https://doi.org/10.1044/2019_JSLHR-L-18-0131
3. Estes, K. G.; Evans, J. L.; Alibali, M. W.; & Saffran, J. R. Can infants map meaning to newly segmented words? Statistical segmentation and word learning. *Psychol. Sci.* **2007**, *18*(3) 254–260. <https://doi.org/10.1111/j.1467-9280.2007.01885.x>
4. Hay, J. F.; Pelucchi, B.; Estes, K. G.; & Saffran, J. R. Linking sounds to meanings: Infant statistical learning in a natural language. *Cogn. Psychol.* **2011**, *63*, 93–106. <https://doi.org/10.1016/j.cogpsych.2011.06.002>
5. Mintz, T. H. Frequent frames as a cue for grammatical categories in child directed speech. *Cognition* **2003** *90*(1), 91–117. [https://doi.org/10.1016/S0010-0277\(03\)00140-9](https://doi.org/10.1016/S0010-0277(03)00140-9)
6. Monaghan, P.; Christiansen, M. H.; Chater, N. The phonological-distributional coherence hypothesis: Cross-linguistic evidence in language acquisition. *Cogn. Psychol.* **2007**, *55*(4), 259–305. <https://doi.org/10.1016/j.cogpsych.2006.12.001>

7. Gómez, R.; Maye, J. The developmental trajectory of nonadjacent dependency learning. *Infancy* **2005**, *7*, 183-206. https://doi.org/10.1207/s15327078in0702_4
8. Hsu, H.; Tomblin, J.; Christiansen, M. Impaired statistical learning of non-adjacent dependencies in adolescents with specific language impairment. *Front. Psychol.* **2004**, *5*, 175. <https://doi.org/10.3389/fpsyg.2014.00175>
9. Kidd, E. Implicit statistical learning is directly associated with the acquisition of syntax. *Dev. Psychol.* **2012**, *48*(1), 171. <https://doi.org/10.1037/a0025405>
10. Newport, E. L.; Aslin, R. N. Learning at a distance I. Statistical learning of non-adjacent dependencies. *Cogn. Psychol.* **2004**, *48*(2), 127-162. [https://doi.org/10.1016/S0010-0285\(03\)00128-2](https://doi.org/10.1016/S0010-0285(03)00128-2)
11. Arciuli, J.; Simpson, I. C. Statistical learning is related to reading ability in children and adults. *Cogn. Sci.* **2012**, *36*, 286-304. <https://doi.org/10.1111/j.1551-6709.2011.01200.x>
12. Lages, A.; Oliveira, H. M.; Arantes, J.; Gutiérrez-Domínguez, F. J.; Soares, A. P. Drawing the links between statistical learning and children's spoken and written language skills. In *International Handbook of Clinical Psychology*, 1st ed; Buela-Casal G., Ed.; Thomson Reuters Editorial, 2022. ISBN: 978-84-1390-874-8
13. Sawi, O. M.; Rueckl, J. Reading and the neurocognitive bases of statistical learning. *Sci. Stud. Read.* **2019**, *23*(1), 8-23. <https://doi.org/10.1080/10888438.2018.1457681>
14. Spencer, M.; Kaschak, M. P.; Jones, J. L.; Lonigan, C. J. Statistical learning is related to early literacy-related skills. *Read. Writ.* **2015**, *28*(4), 467-490. <https://doi.org/10.1007/s11145-014-9533-0>
15. Siegelman, N. Statistical learning abilities and their relation to language. *Lang. Linguist. Compass* **2020**, *14*(3), e12365. <https://doi.org/10.1111/lnc3.12365>
16. Johnson, E. K. Bootstrapping language: Are infant statisticians up to the job? In *Statistical learning and language acquisition*; Rebuschat, P., Williams, J., Eds.; Mouton de Gruyter, 2012, pp. 55-89. <https://doi.org/10.1515/9781934078242.55>
17. Yang, C. D. Universal grammar, statistics or both? *Trends Cogn. Sci.* **2004**, *8*, 451-456. <https://doi.org/10.1016/j.tics.2004.08.00>
18. Batterink, L. J.; Paller, K. A. Online neural monitoring of statistical learning. *Cortex* **2017**, *90*, 31-45. <https://doi.org/10.1016/j.cortex.2017.02.004>
19. Batterink, L. J.; Reber, P. J.; Paller, K. A. Functional differences between statistical learning with and without explicit training. *Learn. Mem.* **2015**, *22*(11), 544-556. <https://doi.org/10.1101/lm.037986.114>
20. Batterink, L.; Reber, P. J.; Neville, H.; Paller, K. A. Implicit and explicit contributions to statistical learning. *J. Mem. Lang.* **2005**, *83*, 62-78. <https://doi.org/10.1016/j.jml.2015.04.004>
21. Cunillera, T.; Càmarà, E.; Toro, J. M.; Marco-Pallares, J.; Sebastián-Galles, N.; Ortiz, H. Time course and functional neuroanatomy of speech segmentation in adults. *Neuroimage* **2009**, *48*, 541-553. <https://doi.org/10.1016/j.neuroimage.2009.06.069>
22. Sanders, L. D.; Ameral, V.; Sayles, K. Event-related potentials index segmentation of nonsense sounds. *Neuropsychologia* **2009**, *47*, 1183-1186. <https://doi.org/10.1016/j.neuropsychologia.2008.11.005>
23. Soares, A. P.; Gutiérrez-Domínguez, F.; Vasconcelos, M.; Oliveira, H.M.; Tomé, D.; Jiménez, L. Not all words are equally acquired: Transitional probabilities and instructions affect the electrophysiological correlates of statistical learning. *Front. Hum. Neurosci.* **2020**, *14*, 577991. <https://doi.org/10.3389/fnhum.2020.577991>
24. Zipf, G. *The Psychobiology of Language*; Routledge: London, England, 1936.
25. Piantadosi, S. T. Zipf's word frequency law in natural language: A critical review and future directions. *Psychon. Bull. Rev.* **2014**, *21*(5), 1112-1130.
26. Soares, A. P.; Iriarte, A.; Almeida, J. J.; Simões, A.; Costa, A.; França, P.; Machado, J.; Comesaña, M. Procura-PALavras (P-PAL): Uma nova medida de frequência lexical do português europeu contemporâneo. *Psicol. Reflex. Crit.* **2014**, *27*(1), 1-14. <http://dx.doi.org/10.1590/S0102-79722014000100013>
27. Soares, A. P.; Lages, A.; Silva, A.; Comesaña, M.; Sousa, I.; Pinheiro, A. P.; Perea, M. Psycholinguistics variables in the visual-word recognition and pronunciation of European Portuguese words: A megastudy approach. *Lang. Cogn. Neurosci.* **2019**, *34*(6), 689-719. <https://doi.org/10.1080/23273798.2019.1578395>

28. Campos, A. D.; Oliveira, H. M.; Soares, A. P. The role of syllables in intermediate-depth stress-timed languages: Masked priming evidence in European Portuguese. *Read. Writ.* **2018**, *31*, 1209-1229. <https://doi.org/10.1007/s11145-018-9835-8>
29. Thiessen, E. D.; Kronstein, A. T.; Hufnagle, D. G. The extraction and integration framework: A two-process account of statistical learning. *Psychol. Bull.* **2013**, *139*, 792–814. <https://doi.org/10.1037/a0030801>
30. Hasson, U. The neurobiology of uncertainty: implications for statistical learning. *Philos. Trans. R. Soc. B* **2017**, *372*(1711), 20160048. <https://doi.org/10.1098/rstb.2016.0048>
31. Stärk, K.; Kidd, E.; Frost, R. L. The effect of children's prior knowledge and language abilities on their statistical learning. *Appl. Psycholinguist.* **2022**, *43*(5), 1045-1071.
32. Siegelman, N.; Bogaerts, L.; Elazar, A.; Arciuli, J.; Frost, R. Linguistic entrenchment: Prior knowledge impacts statistical learning performance. *Cognition* **2018**, *177*, 198–213. <https://doi.org/10.1016/j.cognition.2018.04.011>
33. Cunillera, T.; Laine, M.; Càmarà, E.; Rodríguez-Fornells, A. Bridging the gap between speech segmentation and word-to-world mappings: Evidence from an audiovisual statistical learning task. *J. Mem. Lang.* **2010**, *63*(3), 295-305. <https://doi.org/10.1016/j.jml.2010.05.003>
34. Bogaerts, L.; Siegelman, N.; Frost, R. Splitting the variance of statistical learning performance: a parametric investigation of exposure duration and transitional probabilities. *Psychol. Bull.* **2016**, *23*, 1250–1256. <https://doi.org/10.3758/s13423-015-0996-z>
35. Gutiérrez-Domínguez, F. J.; Lages, A.; Oliveira, H. M.; Soares, A. P. Neural signature of statistical learning: Proposed signs of typical/atypical language functioning from EEG time-frequency analysis. In *International Handbook of Clinical Psychology*, 1st ed; Buéla-Casal G., Ed.; Thomson Reuters Editorial, 2022. ISBN: 978-84-1390-874-8
36. Lavi-Rotbain, O.; Arnon, I. The learnability consequences of Zipfian distributions in language. *Cognition* **2022**, *223*, 105038. <https://doi.org/10.1016/j.cognition.2022.105038>
37. Siegelman, N.; Bogaerts, L.; Frost, R. Measuring individual differences in statistical learning: current pitfalls and possible solutions. *Behav. Res. Methods* **2017**, *49*, 418–432. <https://doi.org/10.3758/s13428-016-0719-z>
38. Bogaerts, L.; Siegelman, N.; Christiansen, M. H.; Frost, R. Is there such a thing as a 'good statistical learner'? *Trends Cogn. Sci.* **2022**, *26*(1), 25-37. <https://doi.org/10.1016/j.tics.2021.10.012>
39. Soares, A. P.; Gutiérrez-Domínguez, F. J.; Lages, A.; Oliveira, H. M.; Vasconcelos, M.; Jiménez, L. Learning Words While Listening to Syllables: Electrophysiological Correlates of Statistical Learning in Children and Adults. *Front. Hum. Neurosci.* **2022**, *16*, 805723. <https://doi.org/10.3389/fnhum.2022.805723>
40. Soares, A. P.; Gutiérrez-Domínguez, F.; Oliveira, H. M.; Lages, A.; Guerra, N.; Pereira, A. R.; Tomé, D.; Lousada, M. Explicit instructions do not enhance auditory statistical learning in children with developmental language disorder: Evidence from ERPs. *Front. Psychol.* **2022**, *13*, 905762. <https://doi.org/10.3389/fpsyg.2022.905762>
41. Soares, A. P.; Lages, A.; Oliveira, H. M.; Gutiérrez-Domínguez, F.J. Extracting word-like units when two concurrent regularities collide: Electrophysiological evidence. In *Proceedings of 12th International Conference of Experimental Linguistics*, Athens, Greece, 11-13 October, 2021; Botinis, A., Ed.; ExLing Society, Athens, Greece, 2021. <https://doi.org/10.36505/ExLing-2021/12>
42. Soares, A. P.; Lages, A.; Oliveira, H. M.; Gutiérrez-Domínguez, F. J. Can explicit instructions enhance auditory statistical learning in children with Developmental Language Disorder? In *International Handbook of Clinical Psychology*, 1st ed; Buéla-Casal G., Ed.; Thomson Reuters Editorial, 2022. ISBN: 978-84-1390-874-8
43. Bishop, D. V. M.; Snowling, M. J.; Thompson, P. A.; Greenhalgh, T.; The CATALISE Consortium Phase 2 of CATALISE: a multinational and multidisciplinary Delphi consensus study of problems with language development: Terminology. *J. Child Psychol. Psychiatry* **2017**, *58*(10), 1068-1080. <https://doi.org/10.1111/jcpp.12721>
44. Soares, A. P.; Lousada, M.; Ramalho, M. Perturbação do desenvolvimento da linguagem: Terminologia, caracterização e implicações para os processos de alfabetização [Language development disorder: Terminology, characterization and implications for literacy]. In *Alfabetização Baseada na Ciência*; Alves, R. A., Leite, I., Eds; CAPES, 2021, pp. 441-471, ISBN: 978-65-87026-86-2

45. Soares, A. P.; França, T.; Gutiérrez-Domínguez, F.; Sousa, I. & Oliveira, H. M. As trials goes by: Effects of 2-AFC item repetition on SL performance for high- and low-TP 'Words' under implicit and explicit conditions. *Can. J. Exp. Psychol.* **2022**, 77(1), 57-72. <https://doi.org/10.1037/cep0000290>
46. Abila, D.; Katahira, K.; Okanoya, K. On-line assessment of statistical learning by event-related potentials. *J. Cogn. Neurosci.* **2008**, 20, 952-964. <https://doi.org/10.1162/jocn.2008.20058>
47. Lukács, Á.; Lukics, K. S.; Dobó, D. Online Statistical Learning in Developmental Language Disorder. *Front. Hum. Neurosci.* **2021** 15, 715818. <https://doi.org/10.3389/fnhum.2021.715818>
48. Raviv, L.; Arnon, I. The developmental trajectory of children's auditory and visual statistical learning abilities: modality-based differences in the effect of age. *Dev. Sci.* **2018** 21, e12593. <https://doi.org/10.1111/desc.12593>
49. van Witteloostuijn, M.; Lammertink, I.; Boersma, P.; Wijnen, F.; Rispens, J. Assessing visual statistical learning in early-school-aged children: The usefulness of an online reaction time measure. *Front. Psychol.* **2019** 10, 2051. <https://doi.org/10.3389/fpsyg.2019.02051>
50. Soares, A. P.; Silva, R.; Faria, F.; Mendes, H. O.; Jiménez, L. Literacy effects on artificial grammar learning (AGL) with letters and colors: Evidence from preschool and primary school children. *Lang. Cogn.* **2021**, 13(4), 534 - 561. <https://doi.org/10.1017/langcog.2021.12>
51. Daikoku, T.; Yatomi, Y.; Yumoto, M. Statistical learning of an auditory sequence and reorganization of acquired knowledge: A time course of word segmentation and ordering. *Neuropsychologia* **2017**, 95, 1-10. <https://doi.org/10.1016/j.neuropsychologia.2016.12.006>
52. Siegelman, N.; Bogaerts, L.; Frost, R. What Determines Visual Statistical Learning Performance? Insights From Information Theory. *Cogn. Sci.* **2019**, 43(12). <https://doi.org/10.1111/cogs.12556>
53. Shannon, C. Claude Shannon. *Information Theory* **1948**, 3, 224.
54. Daikoku, T.; Yumoto, M. Order of statistical learning depends on perceptive uncertainty. *Curr. Res. Neurobiol.* **2023**, 4, 100080.
55. Aslin, R. N.; Newport, E. L. Statistical learning: From acquiring specific items to forming general rules. *Curr. Dir. Psychol. Sci.* **2012**, 21(3), 170–176. <https://doi.org/10.1177/0963721412436806>
56. Cubit, L. S.; Canale, R.; Handsman, R.; Kidd, C.; Bennetto, L. Visual Attention Preference for Intermediate Predictability in Young Children. *Child Dev.* **2021**, 92(2), 691-703. <https://doi.org/10.1111/cdev.13536>
57. Fiser, J.; Aslin, R. N. Statistical learning of higher-order temporal structure from visual shape sequences. *J. Exp. Psychol. Learn. Mem. Cogn.* **2002**, 28(3), 458–467. <https://doi.org/10.1037/0278-7393.28.3.458>
58. Kidd, C.; Piantadosi, S. T.; Aslin, R. N. The Goldilocks Effect: Human Infants Allocate Attention to Visual Sequences That Are Neither Too Simple Nor Too Complex. *PloS ONE* **2012**, 7(5), e36399.. <https://doi.org/10.1371/journal.pone.0036399>
59. Siegelman, N.; Bogaerts, L.; Frost, R. Measuring individual differences in statistical learning: Current pitfalls and possible solutions. *Behav. Res. Methods* **2017**, 49(2), 418–432. <https://doi.org/10.3758/s13428-016-0719-z>
60. Wade, S.; Kidd, C. The role of prior knowledge and curiosity in learning. *Psychon. Bull. Rev.* **2019** 26(4), 1377–1387. <https://doi.org/10.3758/s13423-019-01598-6>
61. Heinks-Maldonado, T. H.; Mathalon, D. H.; Gray, M.; Ford, J. M. Fine-tuning of auditory cortex during speech production. *Psychophysiology* **2005**, 42, 180–190. <https://doi.org/10.1111/j.1469-8986.2005.00272.x>
62. Siegelman, N.; Bogaerts, L.; Christiansen, M. H.; Frost, R. Towards a theory of individual differences in statistical learning. *Philos. Trans. R. Soc. B* **2017**, 372(1711), 20160059. <https://doi.org/10.1098/rstb.2016.0059>
63. Faul, F.; Erdfelder, E.; Buchner, A.; Lang, A. G. Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behav. Res. Methods* **2009**, 41, 1149–1160 <https://doi.org/10.3758/BRM.41.4.1149>
64. Oldfield, R. C. The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia* **1971**, 9, 97–113. [https://doi.org/10.1016/0028-3932\(71\)90067-4](https://doi.org/10.1016/0028-3932(71)90067-4)
65. Espirito-Santo, H.; Pires, C. F.; Garcia, I. Q.; Daniel, F.; Silva, A. G. D.; Fazio, R. L. Preliminary validation of the Portuguese Edinburgh Handedness Inventory in an adult sample. *Appl. Neuropsychol. Adult* **2017**, 24, 275-287. <https://doi.org/10.1080/23279095.2017.1290636>

66. De Diego Balaguer, R.; Toro, J. M.; Rodríguez-Fornells, A.; Bachoud-Lévi, A. Different neurophysiological mechanisms underlying word and rule extraction from speech. *PLoS ONE* **2007**, *2*, e1175. <https://doi.org/10.1371/journal.pone.0001175>
67. Cunillera, T.; Toro, J. M.; Sebastián-Gallés, N.; Rodríguez-Fornells, A. The effects of stress and statistical cues on continuous speech segmentation: an event-related brain potential study. *Brain Res.* **2006**, *1123*(1), 168–178. <https://doi.org/10.1016/j.brainres.2006.09.046>
68. François, C.; Teixidó, M.; Takerkart, S.; Agut, T.; Bosch, L.; Rodríguez-Fornells, A. Enhanced Neonatal Brain Responses To Sung Streams Predict Vocabulary Outcomes By Age 18 Months. *Sci. Rep.* **2017**, *7*, 1245. <https://doi.org/10.1038/s41598-017-12798-2>
69. Pierce, L. J.; Tague, E. C.; Nelson III, C. A. Maternal stress predicts neural responses during auditory statistical learning in 26-month-old children: an event-related potential study. *Cognition* **2021**, *213*, 104600. <https://doi.org/10.1016/j.cognition.2021.104600>
70. Eakin, D. K.; Smith, R. Retroactive interference effects in implicit memory. *J. Exp. Psychol. Learn. Mem. Cogn.* **2012**, *38*(5), 1419–1424. <https://doi.org/10.1037/a0027208>
71. Jiménez, L.; Vaquero, J. M. M.; Lupiáñez, J. Qualitative differences between implicit and explicit sequence learning. *J. Exp. Psychol. Learn. Mem. Cogn.* **2006**, *32*(3), 475–490. <https://doi.org/10.1037/0278-7393.32.3.475>
72. Lustig, C.; Hasher, L. Implicit memory is not immune to interference. *Psychol. Bull.* **2001**, *127*(5), 618–628. <https://doi.org/10.1037/0033-2909.127.5.618>
73. Vaquero, J. M. M.; Lupianez, J.; Jiménez, L. Asymmetrical effects of control on the expression of implicit sequence learning. *Psychol. Res.* **2019**, *1*, 1–15. <https://doi.org/10.1007/s00426-019-01222-1>

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.