Speech, pitch and rhythmic anticipatory abilities in monolinguals and bilinguals

# 1. Introduction

Some linguistic theories consider language to be an independent module in the human cognitive domain (e.g., Chomsky, 1986; Fodor, 1983). In other models, language is an independent cognitive capacity associated with other domains (e.g., Hasson, Egidi, Marelli, & Willems, 2018), but the nature of this association is unknown. In the field of language anticipation (henceforth, linguistic anticipation and linguistic prediction), many scholars advocate for a more global view where language is not independent and its correct functioning depends on the collaboration of larger non-linguistic cognitive capacities (e.g., Ryskin, Levy, & Fedorenko, 2020). In a predictive processing view of language, a more general point of view does not sound implausible. There is abundant evidence that all species make predictions (e.g., animals, Alcaro & Panksepp, 2011; Wilson & Lindstrom, 2011; plants, Appel & Cocroft, 2014). And humans do not only have the unique ability to generate linguistic predictions (e.g, anticipating the end of words) but additionally make non-linguistic predictions (e.g., while driving, Morando, Victor, & Dozza, 2016; Stahl, Donmez, & Jamieson, 2016; or while playing sports, Nakamoto & Mori, 2012; Wright, Bishop, Jackson, & Abernethy, 2010).

In my dissertation, I investigate whether there is an association between a person’s ability to make non-linguistic and linguistic predictions. Finding an association between linguistic and non-linguistic domains and the nature of it, if there is one, is essential given that anticipation is a basic cognitive action deployed continuously to facilitate information processing in all possible domains. Understanding the association among domains would clarify how different types of cognitive abilities are related, and how different types of cognitive experiences affect encoding and processing of other theoretically unrelated domains, like for example music and driving.

To address the possible relationship between domains, I examine auditory (pitch, rhythm) and visual-spatial (object in motion) non-language anticipation and language anticipation (a word’s suffix) in adult Spanish monolinguals and English and Mandarin Chinese intermediate and advanced learners of Spanish. Auditory anticipation was selected because numerous studies show that processing in language and other auditory domains such as music are interrelated. For pitch, musicians of a non-tonal L1 are more sensitive to linguistic tone differences than non-musicians of the same L1 (Chua & Brunt, 2014), and non-musician speakers of tonal L1s perform musical tasks better than non-musician speakers of non-tonal L1s (e.g., Bidelman, Hutka, & Moreno, 2013; Ngo, Vu, & Strybel, 2016). However, no study has examined whether this extrapolated ability can also be used for prediction or only in perception. For rhythm, rhythm perception and production abilities are positively correlated with foreign prosody perception, and in particular with lexical stress (Cason et al., 2019). In contrast to pitch, prediction studies on rhythm have shown that rhythm anticipation is linked to linguistic abilities. For instance, typically developing children anticipate upcoming beats better than dyslexic children (Persici, Stucchi, & Arosio, 2019). Finally, for visual-spatial processing, some scholars propose that children’s dyslexia is caused by deficits in visual-spatial attention rather than to limited phonological processing (Vidyasagar & Pammer, 2010). However, it is not known how these deficits in visual-spatial processing are related to anticipation of linguistic and visual-spatial information.

In addition to exploring the role of anticipatory experience (association between non-linguistic and linguistic prediction), I investigate the role of language experience (L1 extrapolation; L2 proficiency) and working memory (WM) in adult second language acquisition (SLA). In a usage-based framework, language experience is important because it shapes grammar constantly (Bybee, 2006). Language experience shifts a speaker’s attention to and inhibition of input depending on whether the new input provides information that can alter the representations created or not. Thus, language experience keeps affecting one’s ability to make predictions in any language. The inhibition effect is called blocking (Kamin, 1969) and consists of cues predicting a certain outcome not being encoded as cues because another cue is already encoded as a predictor of that outcome. That is, the speaker has learned to pay attention to that reliable association (learned attention, Mackintosh, 1975). In the usage-based framework, a speaker in the L2 should be able to resort to that association for successful L2 prediction. However, cue-outcome associations do not always overlap completely, so in the case of partial overlap, it is unknown whether the similarity between the associations or the discrepancy are stronger in impacting language processing.

The ability to make cue-outcome associations and use them in prediction may well be conditioned by WM. WM has been shown to contribute to language anticipation sometimes (Huettig & Mani, 2016) but not others (Otten & Van Berkum, 2009). The main structures selected to research the influence of WM on anticipation have been morphosyntactic. The function it plays may, however, be the responsible for the differences in WM findings. Huettig and Mani (2016) researched anticipation of morphosyntax as an outcome in L1 Dutch and did find WM differences, while Otten and Van Berkum (2009) used morphosyntax as the cue and found no effects of WM. It is, however, not possible so far to draw many conclusions from the available results on the role of WM on linguistic anticipation as they are scant. The grounds set by the differences in the use of morphosyntax in language anticipation in Huettig and Mani (2016) and Otten and Van Berkum (2009)’s studies may be a source of findings on the role of WM on language processing and anticipation if that line of research is further pursued.  
Indeed, morphosyntactic anticipation in general has been very productive in providing insight into the mechanisms of linguistic prediction. Overall, morphosyntactic information such as gender can be predicted in an L1 (Kamide, Altmann, & Haywood, 2003) and already in an L2 when the L1 has a gender system as well at beginner stages (Dussias, Kroff, Tamargo, & Gerfen, 2013). Other types of morphosyntactic information such as noun number (Roll, Horne, & Lindgren, 2010) and tense (Sagarra & Casillas, 2018) can also be predicted based on phonological cues.

To investigate the influence of language experience on morphosyntactic prediction based on phonological cues, I compare English and Mandarin Chinese learners of Spanish. Relevant to my dissertation, Mandarin, English and Spanish differ in the presence of tones (present in Mandarin, absent in English and Spanish), lexical stress (present in English and Spanish, absent in Mandarin, although Mandarin tones share the main acoustic correlate with Spanish lexical stress), and morphology (null in Chinese, poor in English, rich in Spanish).

Studies on learning experience have revealed that adults can anticipate linguistic information in their L1 (e.g., Roll, Söderström, & Horne, 2013) and might do so in their L2 (e.g., Dussias et al., 2013), although only extensive active practice results in L2 performance comparable to L1 performance (Lozano-Argüelles, Sagarra, & Casillas, 2019). L2 processing and anticipation capabilities are largely constrained by the speaker’s proficiency (e.g., Sagarra & Casillas, 2018) and cross-linguistic effects from the L1 (e.g., Dupoux, Sebastián-Gallés, Navarrete, & Peperkamp, 2008), although how and when the L1 affects L2 processing and anticipation is not clear.

Studies on the association of auditory abilities have demonstrated that processing and comprehension abilities are not only extrapolated across languages (e.g., Ghaffarvand Mokari & Werner, 2018), but also in terms of language and pitch and rhythm abilities (e.g., Tsukada, Xu, & Rattanasone, 2015; Cason et al., 2019, respectively). However, no study has explored the possibility that this extrapolation also reveals itself in anticipation, even though anticipation is present in other auditory domains such as in harmonic, rhythmic, or melodic processing (Patel & Iversen, 2014; Schellenberg, 1996; Tillmann, 2012). Similarly, a relationship has been found between speech processing abilities and visual-spatial abilities (Phillips, Jarrold, Baddeley, Grant, & Karmiloff-Smith, 2004), but no research on how this relationship extends to anticipation performance has been undertaken.

This project aims to study cross-linguistic associations over different levels of L2 proficiency, to examine the ability to associate acoustic knowledge to predict non-linguistic information (pitch, rhythm, space), and to explore the role of cognitive individual differences, i.e. WM, in each of these cases. The project is hence split into three studies: transfer within domain (L1>L2), within modality but between domains (speech>music), and between modaliltieses and domains (speech>vision). By completing these studies, this project will inform models of phonological transfer as well as models of WM, by establishing how different cognitive domains are interrelated.

# 2. Background

## 2.1. Anticipation

Anticipation facilitates the processing of information that we receive from the world: we anticipate the denouement in movies and books, we anticipate lyrics and instrumental parts in songs, and we anticipate what we read and hear in speech. In non-linguistic auditory domains, like music, we anticipate the melody through pitch and rhythm. In speech anticipation, we use phonological, morphosyntactic, and contextual cues to predict incoming linguistic information. Predictive processing in language is especially efficient in the L1, but not so much in the L2. Linguistic anticipation of speech in the L2 may be mediated by factors such as L1 transfer, L2 proficiency, or working memory (WM), but also by other non-linguistic auditory factors such as rhythmic or melodic abilities.

**2.1.1. Linguistic anticipation**

In a predictive processing view of language, linguistic outcomes are predicted by means of cues already available (e.g., Federmeier, 2007; Wicha, Bates, Moreno, & Kutas, 2003). Language processing is thus not cognitively taxing and information is not lost in a noisy and fastly changing environment (Christiansen & Chater, 2016). The underlying mechanisms of how prediction works are still a matter of debate. Research on populations with reduced or loaded executive resources resulting in underachieved prediction suggests that linguistic anticipation may depend partially on domain-general cognitive resources. For example, the executive functions usually alloted for L1 anticipation are overloaded in L2 learners (Linck, Osthus, Koeth, & Bunting, 2014), so they cannot resort to those functions to generate L2 predictions (e.g., Lew-Williams & Fernald, 2010; Mitsugi & Macwhinney, 2016). In L2 populations, cognitive capacities may also interact with other factors, such as lack of enough linguistic experience (Cuetos, Mitchell, & Corley, 1996) or attention to and inhibition of cross-linguistic effects.

**2.1.1.1. L1 speakers**

Humans anticipate linguistic information at different layers using numerous linguistic cues. Importantly, L1 speakers can use a wide range of phonological cues to generate morphosyntactic predictions. Languages can use phonological cues such as coarticulation (Salverda, Kleinschmidt, & Tanenhaus, 2014), intonation (Nakamura, Arai, & Mazuka, 2012; Weber, Grice, & Crocker, 2006), lexical stress (Correia, Frota, Butler, & Vigário, 2013; Sagarra & Casillas, 2018), pauses between clauses (Hawthorne & Gerken, 2014; Kjelgaard & Speer, 1999), vowel duration (Rehrig, 2017), and tone (Roll & Horne, 2015; Roll, Horne, & Lindgren, 2011) to signal different morphological and morphosyntactic outcomes, like gender, number, person, tense, or aspect.

Suprasegmental phonological information such as lexical tone or lexical stress can cue the generation of the correct anticipation of verbal suffixes and noun choice. Swedish speakers use tonal cues to predict noun number (low tone > singular, *fisken* “fish[SG]”; high tone > plural, *fiskar* “fish[PL],” Roll et al., 2010, 2013; Söderström, Horne, & Roll, 2015) and verb tense (low tone > present, *skrämmer* “I scare”; high tone > past, *skrämde* “I scared,” Söderström, Roll, & Horne, 2012; Roll & Horne, 2015). Similarly, Spanish speakers use lexical stress to predict verbal tense (first syllable stressed > present, *CANta* “he sings”; unstressed > past, *canTÓ* “he sang,” Sagarra & Casillas, 2018) and noun ending (*PRINcipe* “prince” vs. *prinCIPIO* “beginning,” Soto-Faraco, Sebastián-Gallés, & Cutler, 2001). Finally, English natives use vowel duration to predict voice (shorter vowel duration > active: “the girl was pushing the boy”; longer vowel duration > passive: “the girl was pushed by the boy,” Rehrig, 2017).

L1 typical speakers tend to perform predictive language processing with no problems, but there is some individual variability that is still not accounted for. Some studies have tried to explain individual variability through WM, but these studies have not yielded satisfactory results. Sagarra and Casillas (2018) found by means of eye-tracking that WM played no role, or a marginal one in anything, in L1 Spanish speakers’ capacity to anticipate verbal tense based on lexical stress. Similarly but with EEG, Otten and Van Berkum (2009) found no effect of WM on L1 anticipation of the Dutch gender morphosyntactic system, and Pakulak and Neville (2010) found no effect either on expectation violation of determiners or possessives in English. The only study to find positive results was Huettig and Mani (2016)’s, who found also through eye-tracking that the variance between participants in fixating on the correct target noun based on gendered article in Dutch could be accounted for by WM. This study shares with the previous ones methodology, structure under research and population, and L1 experience cannot be used as argument, as gender is used in a daily basis in all language contexts. It is therefore necessary to resort to other factors to explain both the opposite findings and the individual variability. Some of these factors could be different levels of linguistic complexity interacting with behavioral or brain activity performance, the specific contexts in which anticipation takes place, the speaker who uttered the stimuli, or the influence of non-linguistic cognitive factors that might be associated, such as ability to perceive differences in sound quality or nature.

**2.1.1.2. L2 speakers**

If and when L2 speakers are able to anticipate is still an unsolved issue. Particularly in morphological and morphosyntactic anticipation, some studies find that L2 speakers can generate predictions (Dussias et al., 2013), while others have found they cannot (Hopp, 2016), or only in certain situations (Lew-Williams & Fernald, 2010) or with specific levels of proficiency (Sagarra & Casillas, 2018). One factor that could account for the variability in L2 morphological and morphosyntactic anticipation performance is language experience in the form of proficiency and cross-linguistic differences. A second factor that could explain the ability to predict in an L2 or not might be cognitive differences.

Language experience in the form of L1 transfer and L2 proficiency can happen at any layer of linguistic processing, from phonology to discourse structure. Here we review only phonology and morphology. Focusing on morphosyntax, anticipation of morphosyntax in an L2 has been tested primarily with gender (e.g., Hopp, 2013). Other morphosyntactic elements such as number are not predicted as reliably (Marull, 2017), and others, like case, are seemingly not predicted in any context (Hopp, 2015; Mitsugi & Macwhinney, 2016).

As opposed to L1 anticipation, however, morphosyntactic anticipation in the L2 is determined by the listener’s proficiency in the L2 (Sagarra & Casillas, 2018) and their L1. The L1 linguistic system can both help (Dussias et al., 2013) and hinder (Dupoux et al., 2008) L2 processing. As an example, Italian and Spanish both have gender systems, Engish does not. An L1 Italian speaker will predict gender outcomes in L2 Spanish as a beginner due to transfer of gender system knowledge, but it might take an L1 English speaker to reach an upper-intermediate level of proficiency to start to do so (Dussias et al., 2013), since no transfer in this regard is possible. L2 proficiency is nevertheless not all, as the English speaker may also be influenced by the frequency (Grüter & Rohde, 2013) or novelty conditions of the noun (Lew-Williams & Fernald, 2010).

L1 transfer facilitating effects can also be observed in faster and more accurate generation of predictions based on determiners with a similar form in English and Dutch by novice L1 Engish speakers of L2 Dutch (Liburd, 2014). Some accounts argue that lacking the representation of a morphological marking, like gender, in the L1 might not only prevent prediction in an L2 based on that cue, but also hinder it (Hopp, 2016). This proposal fits the inability to make case predictions regardless of the L2 proficiency when the L1 lacks encoded case but the L2 overtly marks it (i.e., English and German, respectively; Hopp, 2015). Sometimes, it is unclear what the cross-linguistic effects are, as L2 anticipation performance might be better explained through L2 proficiency. In cases like intermediate English speakers of L2 Spanish not being able to use determiner number to anticipate number suffixes in a noun but advanced speakers being able to do so (Marull, 2017) may indicate that English speakers are not transferring plural marking knowledge, and instead, it is just increased L2 exposure and practice that allows them to generate the predictions.

When phonology is a cue, L2 experience may enable proficient speakers to use suprasegmental information in a word stem like vowel lengthening (Rehrig, 2017) or lexical tones (e.g., Schremm, Söderström, Horne, & Roll, 2016; Berthelsen, Horne, Brännström, Shtyrov, & Roll, 2018) to predict its suffix when their L1 lacked representation for those phenomena. When the acoustic correlate does have a some studies point in the same direction. Sagarra and Casillas (2018) found that L1 English speakers with advanced L2 Spanish could use lexical stress as a cue to anticipate verbal endings when the stress appeared in CVC syllables but not in CV syllables, and L2 Spanish beginners could not use stress to anticipate verb suffixes in either case. In contrast, other studies suggest that no L2 exposure can aid in easier processing of L2 phonological phenomena. Dupoux et al. (2008) found that speakers of L1 French do not improve over time in their discrimination of lexical stress in L2 Spanish, and hence cannot use it as an anticipatory cue. In comparison to the French Speakers, Cantonese and Mandarin L1 speakers can learn to discriminate lexical stress in an L2 (Chen, 2013; Li, To, & Ng, 2017) and Korean L1 speakers do so to a certain extent (Hualde & Kim, 2015; Lee, Shin, & Garcia, 2019) even though neither of those languages require lexical stress encoding. Whether they can use lexical stress to anticipate other linguistic information has not been tested, although investigations on other suprasegmental cues such as tones (Hed, Schremm, Horne, & Roll, 2019; Schremm et al., 2016) suggest that L2 speakers can learn to use suprasegmental cues in an L2 to anticipate linguistic morphosyntactic information, even if this learned ability does not reach L1-like performance (Perdomo & Kaan, 2019).

Besides language experience, there are other factors that may be conditioning L2 speakers’ ability to generate morphosyntactic predictions, such as WM. WM is the cognitive skill that temporarily stores and processes incoming information so that complex cognitive actions are executed; WM is limited and its capacity varies from person to person (Baddeley, 2007). WM is deployed in all interactions with the world. It affects both linguistic and non-linguistic processing, and the interaction between different types of processing. In connectionist models, WM is domain-free and our processing abilities are determined by domain-general cognitive abilities. Specifically, any capacity is determined by the ability to select which information to pay attention to and which information to inhibit, regardless of the nature of that information. In connectionist models there is no difference between processing and storage. Instead, WM is the activation of part of the long-term memory according to short-term patterns of activations related to domain-specific storages (e.g., Cowan, 2016).

In L2 anticipation, only Sagarra and Casillas (2018) and Lozano-Argüelles (2020) have looked into the influence of verbal WM on prediction in an L2. Sagarra and Casillas (2018) studied anticipation of verbal tense morphosyntax when cued by lexical stress in L1 English learners of L2 Spanish through a visual-world paradigm, and found no effects, either in beginner or advanced learners. Lozano-Argüelles (2020) used the same paradigm, cue-outcome pair, and population and found that WM only mediated L2 morphosyntactic anticipation based on prosodic cues under less demanding conditions. Thus, WM differences, if they have any impact, may only manifest in reduced contexts, such as predicting under excessive cognitive load.

WM also gets deployed in other auditory modalities apart from language. Previous research using a model like Baddeley (2007)’s suggests that children attending music classes score higher in phonological loop component tasks and central executive WM tasks than children who do not receive music training (Roden, Grube, Bongard, & Kreutz, 2014). In rhythmic synchronization, WM has been shown to be a reliable predictor of beat timing calculation in non-musicians (Colley, Keller, & Halpern, 2018). More precisely, Miyake, Onishi, and Poppel (2004) defend that anticipatory synchronization may be influenced by WM in the interstimulus-onset interval of 1800-3600 ms. (Miyake et al., 2004).

Cross-domain research indicates that processing abilities in different areas are associated, and that better WM in one domain can produce effects in another. These effects would be a consequence of the distinct domains sharing span capabilities, while short-term capabilities are more unique to each domain (Kane et al., 2004). In contrast, span capabilities are more structurally diverse than short-term capabilities are. In music, for instance, bilinguals of two languages with an alphabetic system like English and a logosyllabic system like Chinese have more enhanced visuospatial WM than bilinguals of two alphabetic systems, like English and Spanish (Ma, 2016). Observing bilinguals in general, both visual and auditory memory maintenance and manipulation capacities, and bilingual experience interact especially at an L2 intermediate level, while at advanced levels the interaction goes back to a more monolingual-like state (Yang, 2017). The changes in interaction might be due to the progress of L2 acquisition. At intermediate L2 levels, L2 acquisition is in full progress. The L2 has been activated enough to recruit WM to compensate for the lack of linguistic experience, and thus WM is enhanced. At lower levels of proficiency, the brain has not had time yet to create a response to L2 input; and at higher levels, linguistic experience is enough to process, understand and produce the L2, so the supplementation of WM is no longer needed.

Research on bilinguals suggests that they have issues in generating morphosyntactic predictions. L1 transfer and L2 proficiency alone cannot explain the diverging performance pattern when L2 speakers are compared to L1 speakers. Other factors like WM can barely account for individual variability in the L1 or have been scarcely reserached in L2 anticipation. More investigations in this direction is therefore needed. However, it might be necessary to go further up and not only consider WM as regards to language, but where language fits in a speaker’s cognitive capacities. That is, how linguistic anticipation abilities relate to anticipation abilities in other non-linguistic domains.

**2.1.2. Non-linguistic auditory anticipation**

There is abundant evidence supporting a cognitive association between different cognitive domains in an auditory modality. Previous research on music suggests that acoustic features are perceived and sensed fundamentally in the same way regardless of the timbre, and auditory attention predicts fine perception of acoustic cues like AM depth (Carey et al., 2015). These findings lend support to Strait, Kraus, Parbery-Clark, and Ashley (2010)’s hypothesis that fine-grained perception of acoustic signal features is driven by top-down attentional mechanisms. Other research has revealed that a person can be pitch-deaf or beat-deaf, but not necessarily both (Phillips-Silver, Toiviainen, Gosselin, & Peretz, 2013). Investigations like those indicate that rhythm, pitch and other auditory domains are independent but interact with each other through a shared network of domain-general auditory mechanisms that might also underpin speech acquisition, processing and anticipation. Importantly for the current study, behavioral online studies have revealed that musical abilities positively influence the production and perception of linguistic sound structures in L2 learners (Slevc, Davey, Buschkuehl, & Jaeggi, 2016). These results have been replicated in neurocognitive studies in the same population (Marques, Moreno, Luı's Castro, & Besson, 2007), as well as in children (Magne, Schön, & Besson, 2006) and adults (Wong, Skoe, Russo, Dees, & Kraus, 2007). Similarly, other neurocognitive studies reveal that musicians process speech in a more enhanced way and more synchronously to its onset than non-musicians (Musacchia, Sams, Skoe, & Kraus, 2007). Here, I focus on the relationship between language, rhythm and pitch in adult L1 and L2 populations.

*Pitch*

Pitch is the frequency associated to a sound wave that makes it be perceived in a low-high continuum (Klapuri, 2006) and plays a similar role in music and language. In music, pitch affects the note we hear. In language, pitch affects if what we hear is a question (“Coffee?”) or a statement (“Coffee.”). Apart from the phrasal and sentential intonation, pitch also affects lexical tones and predominance against other syllables within a word. However, pitch is not only a common acoustic correlate that speech and music share. Knowledge and perception abilities of pitch in one of the two auditory domains translates in many cases into increased perception abilities in the other domain (Bidelman et al., 2013; Chan & Leung, 2019). This relationship permeates into L1 and L2 perception and processing.

In an L1, L1 speakers of tonal languages detect and discriminate absolute pitch in music faster and better than non-tonal speakers even when they are not musicians (e.g., Chua & Brunt, 2014; Deutsch et al., 2009; Ngo et al., 2016; Tsukada et al., 2015). In fact, they perform as well as non-tonal speaking musically-trained speakers in most pitch and music perception measures (Bidelman et al., 2013). Additionally, musical education contributes to better lexical tone perception even in a tonal language (Tang, Xiong, Zhang, Dong, & Nan, 2016).

In an L2, the influence between music and language is bidirectional, although it does not always have effects. On the one hand, musical knowledge of pitch facilitates perception of L2 tones in musician speakers of non-tonal L1s (Ngo et al., 2016), and non-musician speakers of L1 tonal languages have increased ability in pitch and music perception in general (Bidelman et al., 2013). On the other hand, L1 tone knowledge does not facilitate relative pitch perception, while musical experience does (Ngo et al., 2016), and musical experience does not promote tone-segment connections in a tonal L2 when the L1 was non-tonal, but a tonal L1 does promote L2 tone-segment connection and the creation of a rule-like system of L2 tone information (Chan & Leung, 2019). Music cultural bias cannot account for the inconsistent relationship, as all scales are organized around a shared range of pitch changes (Morrison & Demorest, 2009). Another possible explanation is to explore what cognitive abilities are shared across domains and which ones are domain-specific. A good mechanism for checking this alternative is anticipation.

Humans anticipate melodic information (e.g., Loui & Wessel, 2007; Schellenberg & Habashi, 2015), just like in speech. Melodic expectations are generated based on melodies we have heard before (Schellenberg & Habashi, 2015), in which frequencies’ harmonicity is a cue for consonance (e.g., Bigand, Parncutt, & Lerdahl, 1996; McDermott, Lehr, & Oxenham, 2010; Schön, Gordon, & Besson, 2005), and temporal sound patterns cause prediction of oncoming sounds to be generated (Bailes, Dean, & Pearce, 2013; Pearce & Wiggins, 2012; Rohrmeier & Koelsch, 2012). Moreover, structural priming in non-linguistic pitch sequences leads to easier sentence completion of ambiguous relative clause attachmment (Van de Cavey & Hartsuiker, 2016).

In sum, abilities for speech and pitch in music perception and processing are sometimes related and sometimes they are not. Why the boundary exists and where it lies is still uncertain. Studies on processing and anticipation suggest that there is at least structural cross-domain mechanisms. At deeper levels of expectation generation it is therefore plausible that there is also overlap of processing and anticipatory mechanisms.

*Rhythm*

Rhythm is a pattern of recurrent time intervals that usually occur periodically (Berlyne, 1971), it is possible to predict the start of an interval or the next recurring event based on what has already occurred (Fraisse, 2013; Martin, 1972). Synchronizing to a rhythm therefore involves anticipating when the next interval is coming and acting accordingly. Like pitch, rhythm perception and processing abilities have been linked to L1 and L2 processing.

Regarding adult participants using their L1, brain activity and imaging methods have proven fruitful in obtaining positive evidence for the association of rhythm and speech. Yu et al. (2017) conducted a fMRI study particularly designed to test how different components of music and first language are associated. They compared the results from a group of participants undergoing music training group and a control group in language tests (an animal-word cancellation test and an onset cancellation test) and music tests (an interval test and a rhythmic test). The training group performed better in all tests, and what is more, both accuracy on the animal-word cancellation test and accuracy on the interval test were positively correlated. Similarly, Cason and Schön (2012) demonstrated that on-beat sequence primes yielded faster L1 speech processing than off-beat primes in adults through EEG recordings, and speakers with better synchronizing abilities have more consistent speech-evoked brainstem responses (Tierney & Kraus, 2013).

In relation to rhythm and L2 abilities, sound encoding is favored by rhythmic perception in the L2 experience. Cason et al. (2019) and Bhatara, Yeung, and Nazzi (2015) tested stress placement and rhythmic abilities in L1 French L2 Spanish speakers. The rhythmic production scores of the participants predicted correct placement of nuclear stress in a L2 lexical stress imitation task; and rhythm perception was actually positively associated with L2 learning experience. We also make rhythmic predictions, such that when dancing, we need to orient our temporal attention to the meter of a rhythm in order to process is correctly and adjust our movements (Fujioka, Ross, & Trainor, 2015). Synchronizing to a rhythm further involves to adapt to smaller structures like the rhythm’s beat periods (Burger, Thompson, Luck, Saarikallio, & Toiviainen, 2014). Therefore, in anticipating rhythmic patterns and synchronizing to them we need to perform hierarchial processing, temporal prediction and sequencing (Heard & Lee, 2020), just like we do in language. Furthermore, rhythmic expectations in speech are generated automatically, and rhythmic aptitude is correlated with sensitivity to violations in iambic stress patterns (Magne, Jordan, & Gordon, 2016).

In sum, rhythm and speech processing are connected in both L1 and L2 typical populations. The nature of hierarchical anticipation in both domains and the relationship between rhythmic aptitudes and sensitivity to suprasegmental pattern violations also suggests that the predictive mechanisms might be domain-general. There are different levels at which this common blanket could reveal itself, like structure anticipation, and acoustic realization. It is important to determine how resources for both domains overlap to ascertain how different auditory domains are related in cognition and what effects they can have on each other.

[REFER TO RESULTS IN ARTICLE 1]

# 3. Research questions and hypotheses

This study aims at answering how pitch and rhythmic prediction abilities affect L1 and L2 morphosyntactic prediction abilities, and how the L1 as well as the L2 proficiency and the WM capacities mediate these effects. For that purpose, a group of monolinguals speakers of Spanish, a group of L1 English learners of L2 Spanish and a group of L1 Mandarin learners of L2 Spanish completed several anticipation linguistic and non-linguistic tasks.A population of English native speakers with L2 Spanish was chosen because their L1 and their L2 differ in rhythmic terms due to the different properties of lexical stress in each language. A population of Mandarin speakers stress was chosen because, while rhythmically both Spanish and Mandarin are syllable-timed, tonal native speakers have been shown to perform better at other non-linguistic pitch tasks. The selection of each population thus aims at understanding the relationship between pitch-language and rhythm-language in terms of anticipation abilities.

Lexical stress is the prominence of a syllable that speakers hear relative to the other syllables in the prosodic word (Hualde, 2005). In English and Spanish, lexical stress has no fixed position and is phonologically contrastive at the lexical level in both languages, although this use is much more typical in Spanish than in English. In English it is used predominantly to distinguish heteronyms or pairs of verbs-nouns that have no segmental differences, for instance, “PROduce,” noun vs. to “proDUCE,” verb. In Spanish, lexical stress differentiates all kinds of words and information, such as verbal tense and person (*CANto* ‘I sing’ vs. *canTÓ* ‘s/he sang’), or nouns (*PApa* ‘potato’ vs. *paPÁ* ‘dad’), or nouns from verbs (*TÉRmino* ‘term’ vs. *terMIno* ‘I finish’ vs. *termiNÓ* ‘s/he finished’). In Spanish, the most reliable cues to stress are pitch (F0), duration and intensity (Hualde, 2005; Ortega-Llebaria, 2006; Ortega-Llebaria & Prieto, 2007, 2009). Pitch is higher for stressed syllables and lower for unstressed syllables; regarding intensity, stressed syllables are louder; and lastly, stressed syllables are usually slightly longer. In contrast, the main cues in English are vowel duration and quality (Cooper, Cutler, & Wales, 2002; Cutler, 1986). Typically, unstressed vowels are reduced to [ə].

The different cue weights are also related to the prosodic structure of the language. Vowel reduction is linked to the rhythmic pattern of stressed-timed languages, like English, where the intervals between stressed syllables have similar durations. Spanish, on the contrary, is often described as a syllable-timed language, as syllable duration is quite stable, and vowel quality is steady-state regardless of their tonicity (Colantoni, Steele, Escudero, & Neyra, 2015; Hualde, 2005). In English, the stressed syllable signals a rhythmic unit that can be composed of multiple sub-units until the next stressed syllable, and thus rhythmic unit, arrives. In Spanish, stress is simply part of a syllable, and each syllable is a new rhythmic unit. The rhythmic differences may explain why L1 English speakers encounter difficulties with Spanish lexical stress perception (Face, 2006; Ortega-Llebaria, Gu, & Fan, 2013) and production (Lord, 2007).

The differences and similarities in cue weighting and function can influence lexical stress perception in an L2. For example, Cooper et al. (2002) found that the similar distribution of stress in Dutch and English helped L1 Dutch learners of English transfer their knowledge of lexical stress to process it properly in L2 English. By contrast, German speakers have some difficulty perceiving stress in another lexically free stressed language where lexical stress serves a contrastive function such as Spanish (Schwab & Dellwo, 2016). Regardless of what the difficulties are, L2 speakers of L1s with lexical stress may be able to re-encode lexical stress as a cue for word activation and anticipation in the L2 (Lozano-Argüelles et al., 2019; Sagarra & Casillas, 2018). No study so far has explained how this re-encoding is affected by the L1. If knowledge of the structure and its function can be transferred, or if only acoustic knowledge (e.g., pitch, vowel duration) can be transferred, or if nothing at all is transferred and lexical stress is encoded anew.

Tones are the pitch contour patterns of the voiced part in syllables (Chao, 1968). Only a few languages use tones contrastively at a lexical level. The acoustic correlates for tones vary across languages. While some use also length and/or register (e.g., Cantonese Chinese), for others pitch is the main acoustic correlate in tones (most Mandarin Chinese dialects, Gandour & Fromkin, 1978; Zhu & Wang, 2015). Given that tones in Mandarin Chinese do not cause shorter and longer syllables, Mandarin is described as a syllable-timed language in terms of rhythm (Grabe & Low, 2002; Lin & Wang, 2007; Mok, 2009), just like Spanish. In Mandarin, tones facilitate word recognition but they need the support of segmental information (Hu, Gao, Ma, & Yao, 2012; Malins & Joanisse, 2010; Shen, Hyönä, Wang, Hou, & Zhao, 2020; Wiener & Turnbull, 2016). Since the role of tones is not as pre-eminent as lexical stress in Spanish, Mandarin L1 speakers learning L2 Spanish cannot transfer function knowledge. Instead, they might be able to transfer acoustic knowledge of pitch. If these speakers are able to transfer pitch knowledge, it might help them in learning to use Spanish lexical stress as an anticipatory cue. If there are indeed two types of transfer (acoustic vs. functional), how do both processes differ? And do they entail different cognitive costs?

**RQ. Is linguistic prediction of word morphology associated with auditory prediction in Spanish monolinguals, and intermediate and advanced Chinese and English learners of Spanish? If so, how do WM, L1 transfer and L2 proficiency mediate this relationship?**

**H.** *Pitch*  
In monolingual speakers, better pitch anticipation is not expected to have an effect on individual variability of morphosyntactic anticipation based on lexical stress, as all participants are non-musicians and come from a similar linguistic background. WM is only expected to have an effect on non-linguistic auditory prediction abilities (Roden et al., 2014).

In L2 speakers, Mandarin L1 speakers are expected to anticipate melodic information faster than any of the other populations since they will be extrapolating their pitch knowledge from language to music. However, performance in anticipating melodic information will not be correlated with their L2 anticipation performance. That is, the influence will be unidirectional from language to music. No association between pitch and speech prediction abilities is expected in the English learners. L2 proficiency and pitch anticipatory abilities are not expected to be correlated, as facilitative processing of L2 speech by pitch has only been confirmed with musicians (Slevc & Miyake, 2006). All auditory anticipation scores will be positively correlated with verbal WM scores in the L2 speakers (Yang, 2017).

*Rhythm*

Contrary to pitch, better rhythm anticipation abilities are expected to be associated with individual variability of speech anticipation in monolingual speakers of Spanish. This result would stem from on-beat timings facilitating L1 speech processing (Cason & Schön, 2012), so individuals with better rhythm synchronization abilities would be resorting to their rhythmic abilities to anticipate speech based on prosody more efficiently. Like with pitch, WM is only expected to have an effect on non-linguistic auditory prediction abilities (Colley et al., 2018).

By contrast, rhythmic anticipation abilities are expected to be positively correlated with L2 anticipation performance, as rhythmic abilities are a reliable predictor of lexical stress placement (Cason et al., 2019), and, therefore, of lexical stress comprehension. This influence will be visible at all levels of proficiency and regardless of the L1. That is, better rhythmic abilities will translate into better anticipation performance based on lexical stress when compared to peers around the same proficiency range in both L1s. Like before, L2 proficiency and rhythm anticipatory abilities are not expected to be correlated, because while better rhythm promotes correct L2 perception and production (Cason et al., 2019), anticipation adds another layer of processing difficulty that has been proven to be too complex in the L2 (e.g., Hopp, 2015; Mitsugi & Macwhinney, 2016). All auditory anticipation scores will be positively correlated with verbal WM scores in the L2 speakers (Roden et al., 2014).

# 4. Methods

## 4.1. Participants

Participants proceeded from five different pools: 30 monolinguals speakers of Spanish (MON, females = 20), 30 intermediate L2 Spanish speaekrs with L1 English (IEN, females = 14), 30 advanced L2 Spanish speakers with L1 English (AEN, females = 16), 30 intermediate L2 Spanish speakers with L1 Mandarin (IMA, females = 13), and 30 advanced L2 Spanish speakers with L1 Mandarin (AMA, females = 20). All participants grew up in monolingual regions and homes. Participants of the five groups were aged 18-45 (MON: mean = 26.2, *SD* = 8.82; IEN: mean = 24.9, *SD* = 3.75; AEN: mean = 26.3, *SD* = 4.64; IMA: mean = 24.1, *SD* = 3.97; AMA: mean = 24.8, *SD* = 4.72). Participants were all right-handed. Additionally, speakers were homogenized for verbal WM. Participants had normal to corrected-to-normal vision and no hearing or motor disability. All data were collected in Spain, to avoid possible biases of context in comparing the L2 speakers (Beatty-Martı'nez et al., 2019).

The monolingual speakers were native to a Peninsular variety of Spanish and did not speak any other language fluently. They might have traveled abroad, but not stayed for a long period of time. They all had completed at least high school.

The Mandarin speakers came from Mandarin-speaking regions in China, except for one who was originally from Taiwan. In some of these regions, dialects might be spoken, but the differences consist mainly of lexical choices. Some of the Chinese speakers also spoke English, but at a maximum of intermediate proficiency. They had never received bilingual English-Mandarin education or lived in an English-speaking country. IMA had resided in Spain for 6 months to 11 years, and AMA for 2 months to 16 years.

The English natives came from English-speaking countries (United States 55%, United Kingdom 23%, Canada 7%, New Zealand 7%, Australia 4%, Ireland 4%). The L2 speakers were matched through DELE scores (see section 4.3.1.) in terms of proficiency across levels and they were all late learners of Spanish. That is, all L2 participants had started learning Spanish after 12 years of age (IEN: mean = 22.3, *SD* = 4.32; AEN: mean = 19.9, *SD* = 2.91; IMA: mean = 21.6, *SD* = 4.31; AMA: mean = 20.6, *SD* = 3.40). On average, all L2 groups used Spanish less than 50% of the time in a normal week (self-reported; in percents, IEN: mean = 31.1, *SD* = 18.7; AEN: mean = 37.2, *SD* = 14.5; IMA: mean = 47.4, *SD* = 19.8; AMA: mean = 47.2, *SD* = 22.6). All L2 participants had also completed studies at least through high school. Participants were only allowed to take part in the studies if they had lived in Spain or any other speaking country for at least 2 months. None of the L2 participants had lived in non-monolingual communities in Spain, or in non-Spanish speaking countries. The residing time in Spain range for IEN was 3 months to 5 years and a half, and for AEN 7 months to 10 years.

## 4.2. Materials

Participants completed a series of tasks that allowed to filter them in or out of the studies, to measure their linguistic anticipation abilities (visual-world paradigm), their auditory anticipation abilities in melodies, i.e. pitch, and rhythm, their visuospatial anticipation abilities, their visuospatial WM, and verbal WM.

**4.2.1. Screening tasks**

The first screening task was a proficiency test. The proficiency test was a shortened version (Sagarra & Herschensohn, 2010) of the *Diploma de Español como Lengua Extranjera* (‘Certificate of Spanish as a Foreign Language’, by Instituto Cervantes). The test consisted of 56 questions: 16 on grammar, 10 on vocabulary, and 20 on reading comprehension. The L2 participants were split into two groups for data collection according to their scores on the test. Participants whose scores ranged from 25 to 39 were assigned to the intermediate groups, and participants whose scores were located at or above 40 were assigned to the advanced groups. Only participants scoring within those two ranges were accepted for the study.

Once in the data collection session, participants were interviewed and completed a language background questionnaire, to make sure that they qualified for the study. The background questionnaire contained questions related to their age, handedness, what language they spoke in their household while growing up, how old they were when they started learning Spanish, how long they had lived in Spanish-speaking countries, how much they used each language per week at the time of data collection, other languages they spoke fluently, and what musical, sport and driving abilities they had. Finally, participants also completed a vocabulary test with some of the words from the linguistic anticipation task. This test had 17 screen with 8 words each, and participants had to select the best equivalent in their L1 for the words provided. One of the options provided was three question marks (“???”), meaning “I do not know the word.”

**4.2.2. Anticipation tasks**

**4.2.2.1. Linguistic anticipation task (Visual-world paradigm)**

An eye-tracking visual-world paradigm task assessed the ability to form stress-suffix predictive associations. The eye movements while hearing the sentences were recorded through an EyeLink 1000 Plus desktop mount eye-tracker from SR Research (sampling rate: 1k Hz; spatial resolution was less than .05o; averaged calibration error: .25-.5o). The monitor used to show the target verbs was a BenQ XL2420TE at a resolution of 1920 x 1080 pixels. The rest of the tasks were administered in PsychoPy v3.2. In all tasks, the instructions were given both orally and in writing. The written instructions were in the participant’s L1. Participants read two verbs on a computer screen side by side ( *salta* ‘s/he jumps,’ *saltó* ‘s/he jumped’) and heard a sentence containing one of the two verbs ( *El ladrón saltó la valla* ‘the thief jumped over the fence’). Their task consisted of selecting the verb they had heard as fast as possible by pressing the right- or left-shift key. When they made their selection, a green rectangle appeared around the selected verb. There were 4 practice sentences, 16 experimental sentences, and 80 fillers.

*Visual stimuli*  
 The words that appeared on the screen were presented as images created with PhotoScape X version 3.0.3., set with Helvetica font and the size 300. The size of each image was set at 1920 x 1080 pixels. The background was white and the color of the letters black. The position of the words on each side of the screen was counterbalanced. In the case of the experimental words, half of the present verbs appeared on the right, and the other half on the left. Counterbalancing changed from participant to participant so that a given target would not always appear on the same side.

The experimental words consisted of 32 pairs of verbs corresponding to the 32 experimental sentences. The experimental verbs were regular, had two syllables, had a CVC initial syllable, all belonged to the -ar conjugation, and were conjugated in the 3rd person present and preterit tense. As a reminder, Spanish regular verbs have 2 cues to tense: initial syllable stress and word ending morphology. Regarding stress, stressed initial syllables indicate that the tense is present ( *salta* ‘s/he jumps’), and unstressed initial syllables indicate that the tense is past (preterit, *saltó* ‘s/he jumped’). Regarding morphology, the verb suffix indicates tense, person and aspect.

*Auditory stimuli*

There were 32 sentences in total (at 16 bits, 44.1 kHz, and an intensity of ~75 dB). The sentences were distributed into blocks by means of a Latin square design. There were 8 blocks. Each block contained only two target sentences, one of each condition. The blocks appeared in a randomized order. Within and across the blocks, the sentences were pseudo-randomized to avoid two experimental sentences of the same condition appearing one after the other.

The experimental sentences (*El ladrón saltó la valla* ‘The thief jumped over the fence’) all had identical syntactic structure and length (exactly 5 words). The verb was only preceded by an NP (the subject) to avoid possible biases towards the present or the preterit verb. The object following the verb was also a NP. Half of the experimental sentences contained a present verb and the other half a preterit verb, but a given participant was only exposed to one of the two conditions for a given sentence (e.g., for sentence number 30, a given participant was only exposed to condition 1 and another to condition 2). In total, each participant was exposed to 16 experimental sentences (half with present verbs and half with preterit verbs). There were four practice sentences, one of each type of sentence included in the task.

**4.2.2.2. Pitch anticipation task**

A pitch task measured participants’ ability to predict melodic information based on previously occurring associated pitch changes. This task was inspired by the melody discrimination tasks typically used in music research. This task was divided into a familiarization phase and a testing phase. In the familiarization phase, participants listened to the melodies with which they were going to work in the task. They heard each melody twice and at the same time saw an arrow indicating in which direction the pitch in the melody changed.

In the experimental phase, participants saw two arrows on the screen side by side, one pointing upwards and the other pointing downwards. At the same time, they heard a melody that changed pitch direction after the third tone. Their task consisted of selecting the arrow that matched the pitch change direction as fast as possible by pressing the right- or left-shift key associated with the corresponding arrow. When they made their selection, the arrow selected turned blue. Two practice melodies were used as practice items, and 6 melodies were played as experimental items. The practice items were repeated twice each, and the experimental melodies were repeated 8 times each. The order of appearance of the melodies was pseudorandomized, so the same melody would not play twice in a row.

After the first two practice melodies, feedback was provided for accuracy of response. The second time the practice melodies played, no feedback was provided. No feedback was provided during the experimental phase either.

The melodies were created in pairs, such that each pair started the same, but the fourth note differed. The experimental melodies were played at a tempo of 60 bpm on a scale of A4. The melodies were pairs, such that two melodies started the same and would differ after the third tone. One of the melodies grew higher in pitch, while the other melody grew lower. A higher note than the base always signaled that a still higher note followed, while a lower note always signaled an even lower note that followed. The first changing tone in each melody was meant to act as a cue for participants to make their choice on whether the fifth note would be higher or lower. After the variation in pitch, the melody remained at that newly reached level so as not to confuse participants on which change to judge. The experimental melody pairs were therefore 1) G#-G#-G#-B-C-C and G#-G#-G#-F-E, 2) C-C-C-C#-E-E and C-C-C-C#-E-E, and 3) E-E-E-F-F#-F# and E-E-E-D#-D. The practice melody pair was A-A-A-B-C-C and A-A-A-G-F#-F#. The test was built and run on PsychoPy v3.2. The melodies were created with BeepBox 3.0.8 (<http://www.beepbox.co>) and saved as an mp3.

**4.2.2.3. Rhythmic anticipation task**

A rhythm tapping task adapted from Pagliarini (2016) assessed the participants’ abilities to adapt to a rhythm and thus predict when changes were coming. This task was divided into a familiarization phase and a testing phase for each rhythm. The testing phase of each rhythm consisted of 10 trials of the same rhythm, with no pauses in between. Participants heard the rhythms and one of the beats (the imperative beat) cued them to press a key on the start of the following beat (the tap-along beat). When they pressed the key, a green dot appeared at the center of the screen. Otherwise, the screen was blank. No other feedback was provided at any time. The familiarization phase had the same length as one trial of the testing phase (~5-6 s.). In the familiarization phase, participants simply listened to the rhythm.

There were 2 practice rhythms and 4 experimental rhythms. Each trial in all rhythms consisted of 8 beats. Of these 8 beats, six were context, one was the cue or imperative beat and the remaining one was the tap-along beat. The impeative and the tap-along beats always appeared together, first the imperative beat and then the tap-along beat. Their position in the beat sequence was random for each trial, except they could not appear in the first and last positions. All tones were always pure tones. The base tones and the tap-along beat had a frequency of 440 Hz, and the imperative beat was added a harmonic to make it distinguishable, so its frequency was 880 Hz. All rhythms were at a reference tempo of 80 bpm.

The first practice rhythm contained eight homogeneus beats of 250 ms. at 650 ms. onset-to-onset intervals. The second practice rhythm contained 8 beats of 200 ms. each. The onset-to-onset intervals alternated between 400 and 800 ms, resembling the heartbeat.

In the experimental phase, the first beat rhythm sequence contained a total of eight beats at 750 ms. onset-to-onset intervals. Each beat lasted 200 ms. The second beat rhythm also contained eight beats at 750 ms. onset-to-onset intervals. Four of the beats were strong and the other four were weak. They always appeared in the sequence strong-weak-strong-weak. The intensity of the weak tones was half of the intensity of the strong tones, and the strong tones lasted 200 ms. and the weak ones 100 ms. The third beat rhythm contained four strong and four weak beats in the order strong-weak-strong-weak. The strong tones lasted for 200 ms. and the weak ones for 100 ms. Between a strong and a weak beat the onset-to-onset interval was 990 ms, and between the weak and the strong beats, the onset-to-onset intervals was 510 ms. The last beat rhythm contained equal tones of 200 ms. duration, but the onset-to-onset intervals were 750 ms. ± a random error of 30% of a duration of 750 ms. This sequence was unpredictable and served as base condition. There were two extra beat rhythms that served as practice sequences. The task was created and administered on PsychoPy v3.2.

### 4.2.3. Working memory task

An adapted version of Unsworth et al.’s (2005) Operation Span task (henceforth, OSpan) was used to assess verbal WM span. This task generates independent measures of storage and processing speed. In a single trial, participants heard a word and heard a simple mathematical problem that could be either true or false (e.g., 2 + 2 = 4). At the same time they heard the mathematical problem, they saw the words TRUE and FALSE on the sides of the screen. They had to select as fast as possible the correct word depending on whether what they heard was true or false by pressing the left- or right-shift key corresponding to the side on which their response was. When they selected the word, it turned blue. This process would repeat a certain number of times until a set was complete. At the end of each set, participants were prompted to write down the individual words they had heard before each problem in the same order they had heard them. The whole task was administered in the participant’s L1.

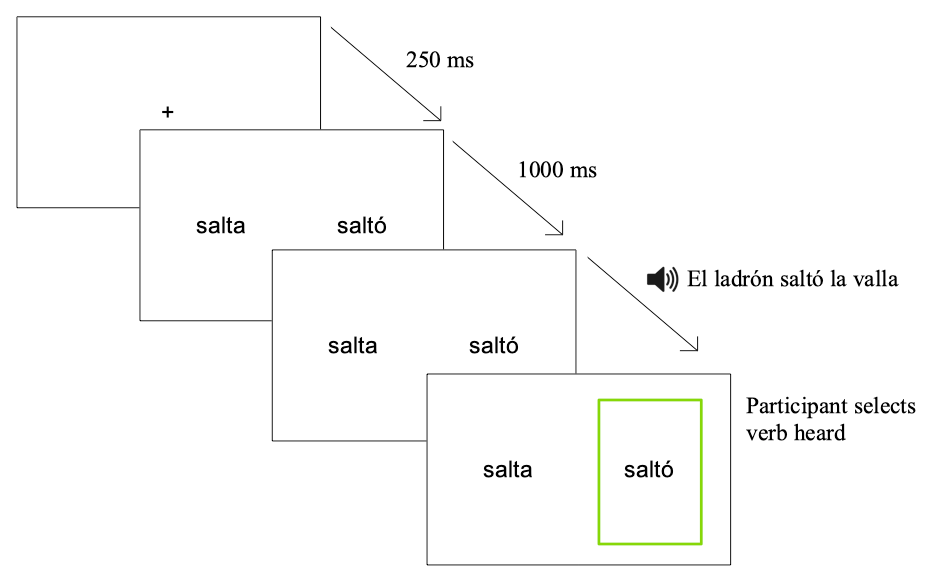
There were two practice trials of 3 words and simple mathematical problems, and three experimental sets of three, four, five and six words and mathematical problems. The words and mathematical problems that participants heard in the sets appeared in a fixed order across participants. The words TRUE and FALSE that participants needed to select in response to the mathematical problems appeared counterbalanced on each side across participants. No feedback was ever provided.

# 4.3. Procedure

Data collection took place in a single session of about 1 hour and 30 minutes to 2 hours depending on the participants. Apart from the participant, only the researcher was present. All interaction happened in Spanish. The room for data collection was isolated from external noise and light. Participants completed the tasks in this order: Spanish proficiency test (only L2 learners; 15-20 minutes), language background questionnaire (10 minutes), eye-tracking task (25 minutes), rhythm task (10 minutes), pitch task (10 minutes), OSpan (15 minutes), and the vocabulary test (10 minutes). First, participants listened to an overview of the tasks and signed the consent form. Participants provided oral responses for the language background questionnaire. They completed the remaining tasks on a computer, using a 24" computer monitor and Sol Republic 1601-32 headphones.

*Linguistic anticipation task*

For the visual-world paradigm, participants rested their head on a chin rest, completed an 11-point grid calibration task, and received task instructions both orally and in writing. Then, they completed the practice trials and were given the opportunity to ask questions. Afterwards, they performed the task. Both the practice and the task trials followed this order. First, participants looked at a fixation sign in the middle of the screen for 250 ms. This fixation sign allowed the researcher to recalibrate manually when necessary. Then, two words appeared in the screen side to side. Once the words had been on the screen for 1000 ms., the sentence started playing and continued until reaching the last word. That is, the sentence did not stop when participants selected the word they heard. Participants were instructed to select the word on the screen they heard on the sentence as fast as possible by pressing the right- or left-shift keys. A green rectangle appeared on the screen around the selected word when participants pressed the key to make their choice. Response recording was set up to be registered only when the press happened after the start of the verb in the sentence. Previous presses where not recorded so the setting forced participants to press again until they saw the green rectangle appear. No feedback was provided, regardless of whether they selected the correct word or not. See Figure 1 for an example of an experimental trial.

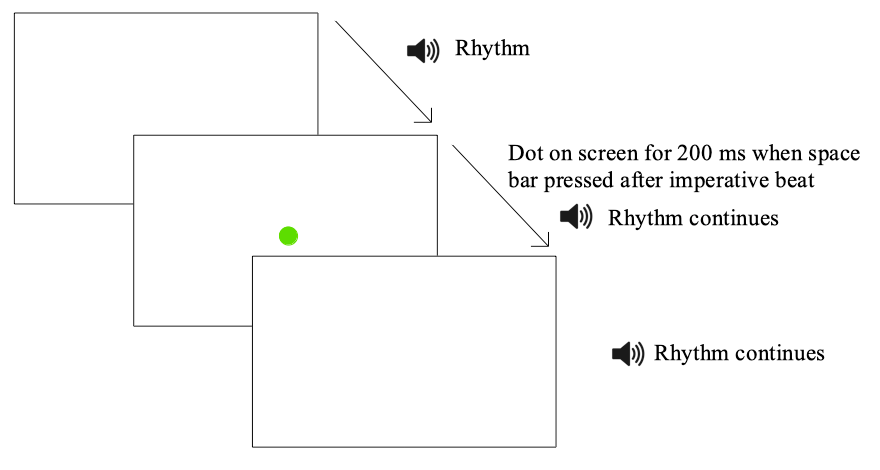


*Figure* *1.* Sample trial of the visual-world paradigm

*Rhythm anticipation task*  
The rhythmic anticipation task consisted of familiarization phases and testing phases. There were 2 rhythms for practice, and 4 experimental rhythms. For each rhythm, participants listened to 6 seconds of the rhythm to become familiarized with it. In the testing phase, one of the beats in the rhythm was replaced by a beep or imperative tone that cued the participants to press the spacebar on the start of the following beat. That is, the order of the sections in the task was:

1. Practice:
   1. Familiarization phase 1
   2. Experimental phase 1
   3. Familiarization phase 2
   4. Experimental phase 2
2. Test:
   1. Familiarization phase 3
   2. Experimental phase 3
   3. Familiarization phase 4
   4. Experimental phase 4
   5. Familiarization phase 5
   6. Experimental phase 5
   7. Familiarization phase 6
   8. Experimental phase 6

There were ten experimental trials for each rhythm in the testing phase, but participants heard no pause between trials. The 10 trials for a single rhythm sounded together in a row as if the trials were all part of the same rhythmic unit. When participants pressed the space bar after the imperative tone, a green dot appeared on the center of the screen. Otherwise, the screen remained blank (see Figure 2 for a trial representation of the rhythm anticipation task).



*Figure* *2.* Sample trial of the rhythm anticipation task

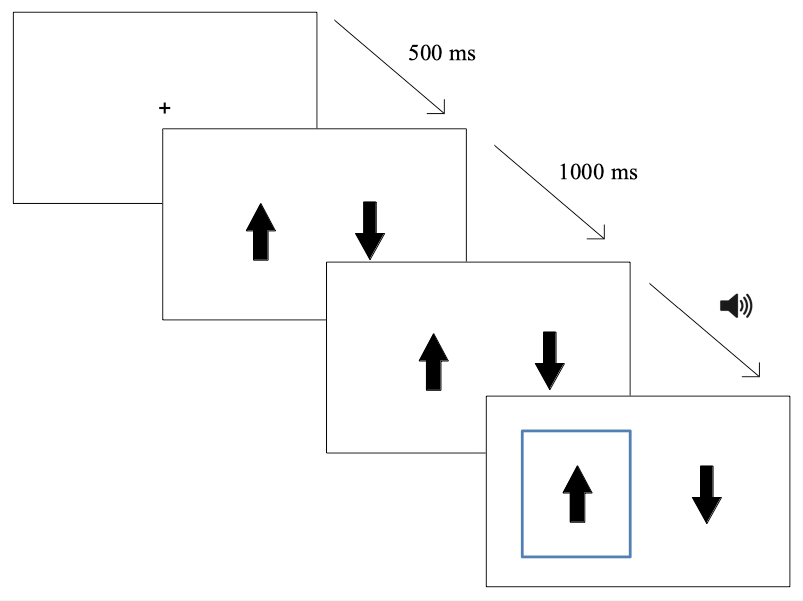
*Pitch anticipation task*  
The pitch anticipation task also contained a familiarization and a testing phase, but in different order. In the pitch anticipation task, participants were familiarized with all the melodies first. The section order was therefore as follows:

1. Familiarization
2. Practice
3. Test

The melodies were four pairs of short melodies. Each pair started the same, but the fourth note differed. Participants were not told about this pairing explicitly. During this familiarization phase, a number and an arrow appeared on the screen. The number (from 1-8) indicated what melody it was but contained no relevant information for the task. The arrow could point upwards or downwards and accompanied the direction of the pitch change.

During the practice and testing phases, an upwards arrow and a downwards arrow appeared on the screen, each on one side. The side where each arrow appeared was counterbalanced across participants. Participants needed to select the arrow corresponding to the pitch change direction as fast as possible by pressing the keys associated with the arrows. The key associated with the arrow on the left was the left-shift key, and the key associated with the arrow on the right was the right-shift key.

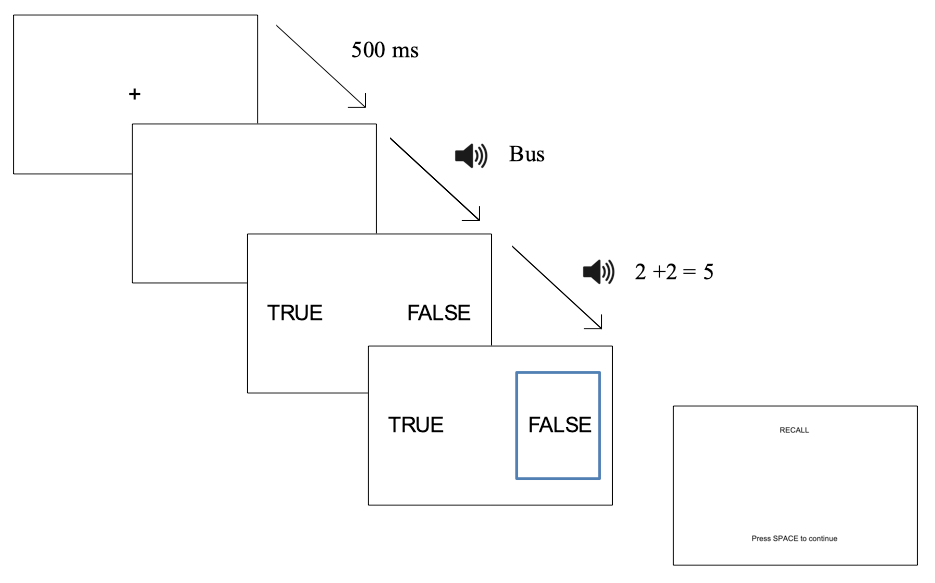
There were four practice trials. In the first two trials, participants received visual feedback on their arrow choice (right vs. wrong), and in the last two practice trials they received no feedback, imitating the format of the experimental trials. The correct sides were counterbalanced across participants. A fixation cross appeared for 250 ms. between trials. A representation of a trial can be seen in Figure 3.



*Figure* *3*. Sample trial of the pitch anticipation task

*Verbal WM task*

The OSpan task was divided into practice and experimental trials. For each trial, participants first heard a word in their L1 and then heard a simple equation, like 2+2=5, also in their L1. During the equation, participants saw the words TRUE or FALSE on each side of the screen, and they had to press the corresponding key (left-shift key for the word on the left, right-shift key for the word on the right) depending on whether the equation was correct or not. After each equation, they saw a fixation point for 500 ms. and another word-equation pair was presented. This process was repeated until the word RECALL or a linguistic equivalent in the L1 of the participant was shown on the screen, at which moment participants had to write on a piece of paper the words they had heard in the order they were presented. The task started with three sets of three trials, then three sets of four trials, and so on until the third set of six trials. Figure 6 shows a sample trial.



*Figure* *4*. Sample trial of the OSpan

*Vocabulary test*

Finally, a vocabulary multiple-choice test evaluated participants’ knowledge of the eye-tracking task words. For each word, they had to select the closest equivalent in their L1 from 8 choices. If they did not know what the word in Spanish meant, they had an option to mark that that was the case. Participants were instructed not to guess the answers if they did not know what a word meant. For the nouns coming fromthe adjective distractor sentences, participants were also required to decided whether the word in Spanish was masculine or feminine. In total, there were 136 items, divided into 17 sections of 8 words. There was no time constraint, and participants could change their answer as long as they did not move on to the following section. The task was set up so they could not go back to previous sections.

# 4.4. Scoring

*Proficiency test*

Each question in the proficiency test was worth one point. Each question answered correctly would thus add one point up. Questions answered incorrectly did not deduct points. Since there were 56 questions, the maximum possible score was 56, and the minimum was 0. Only participants who got at least 25 points were recruited for the study.

*Visual world paradigm*

Participants’ key presses were considered accurate if they pressed the key corresponding to the verb in the sentence, and inaccurate if they pressed the other verb. Only the sentences with accurate responses were included in the statistical analyses. In the statistical analyses, proportion and count of fixations on the verbs over time in the sentence were analyzed. There were several moments in the sentence of special interest, namely, target onset, onset of the first vowel, onset of the code in the first syllable, onset of the second syllable, and offset of the verb. The gaze-fixation data was shifted 200 ms. to account for the time it takes to plan and launch a saccade (e.g., Fischer, 1992; Saslow, 1967), downsampled to 10 ms. and 50 ms. bins and centered at the onset of the last syllable of target items.

*Pitch anticipation task*

Participants’ key presses were considered accurate if they pressed the key corresponding to the direction of the pitch change, and inaccurate if they pressed the key for the other arrow. Only the accurate trials were included in the statistical analyses. Given that participants were primed that a higher pitch to the base indicated that a still higher note followed, and a lower pitch was followed by an even lower tone, the closer the key presses were to the onset of the cueing tone, the better the anticipation was deemed. That is, the shorter the reaction time in pressing the response key, the better. Times were averages for the trials in each melody.

*Rhythm anticipation task*

Following Pagliarini (2016), key presses happening 150 ms. after the beginning of the tap-along beat were discarded for the statistical analyses as they were considered reaction to the second beat. For valid timings, the closer the key presses were to the start of the second beat, the better the anticipation. Times were averaged for the trials in each condition.

*Ospan*

Several measures were taken in this task: for the mathematical problems, if participants pressed TRUE or FALSE correctly, and how long it took them to respond; for the words, if they had written all words and if they had, if they were in the correct order. Given that speakers of different languages have different average spans, processing speed in the mathematical problems was the measure taken to compare WM across L1 groups.

# 4.5. Statistical analyses

Statistical analyses were conducted in R (R Core Team, 2019) version 1.2.5033 using the packages *lme4* (Bates, Mächler, Bolker, & Walker, 2014) and *multcomp* (Hothorn et al., 2016). The empirical logit transformation was applied to binary responses (fixations on target or distractor; Barr, 2008). The bins of 10 ms. were run through independent t-tests to find out whether participants were anticipating the correct verbal tense upon hearing the lexical stress or lack thereof. The bins of 50 ms. were used to 1) model a GLMM to observe how the different factors considered (L1 transfer, L2 proficiency and WM) mediated anticipatory ability; and 2) to model a growth curve analysis (GCA, Mirman, 2016) to analyze the evolution of the gaze fixation pattern over time as suprasegmental and segmental information became available. The GCA models with the time course were implemented by using linear, quadratic, and cubic orthogonal polynomials with the independent variables group and lexical stress. The monolinguals served as baseline. Lexical stress had been sum-coded. By-subject and by-item random effects, and random slopes for the orthogonal time terms were also tested. Nested model comparisons were implemented to assess main effects and interactions.

For the linguistic and tone anticipation tasks, only accurate trials were included. For the rhythm anticipation task, deviation time from target beat onset was averaged across conditions and across the test when the key-presses happened within 150 ms. after the start of the tap-along beat. For the movement anticipation task, deviation time from the ms. when the car should reappear was averaged across conditions and across the test. Reaction time was averaged across conditions and across the test. For the WM tests, number of correct trials out of all trials was the score used as reference. For the OSpan, reaction time in pressing TRUE or FALSE for the mathematical problems was also averaged across trials and used as the measure of processing speed. Proficiency was used to fix groups in advance. The WM scores were analyzed as a covariate.

# References

Alcaro, A., & Panksepp, J. (2011). The seeking mind: Primal neuro-affective substrates for appetitive incentive states and their pathological dynamics in addictions and depression. *Neuroscience & Biobehavioral Reviews*, *35*(9), 1805–1820.

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language*, *38*(4), 419–439.

Altmann, G. T., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, *73*(3), 247–264.

Altmann, G. T., & Kamide, Y. (2007). The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. *Journal of Memory and Language*, *57*(4), 502–518.

Appel, H. M., & Cocroft, R. (2014). Plants respond to leaf vibrations caused by insect herbivore chewing. *Oecologia*, *175*(4), 1257–1266.

Baddeley, A. (2007). *Working memory, thought, and action* (Vol. 45). OUP Oxford.

Bailes, F., Dean, R. T., & Pearce, M. T. (2013). Music cognition as mental time travel. *Scientific Reports*, *3*(1), 1–4.

Barr, D. J. (2008). Analyzing ‘visual world’eyetracking data using multilevel logistic regression. *Journal of Memory and Language*, *59*(4), 457–474.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *arXiv Preprint arXiv:1406.5823*.

Beatty-Martı'nez, A. L., Navarro-Torres, C. A., Dussias, P. E., Bajo, M. T., Guzzardo Tamargo, R. E., & Kroll, J. F. (2019). Interactional context mediates the consequences of bilingualism for language and cognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

Berlyne, D. E. (1971). Aesthetics and psychobiology.

Berthelsen, S. G., Horne, M., Brännström, K. J., Shtyrov, Y., & Roll, M. (2018). Neural processing of morphosyntactic tonal cues in second-language learners. *Journal of Neurolinguistics*, *45*, 60–78.

Bhatara, A., Yeung, H. H., & Nazzi, T. (2015). Foreign language learning in french speakers is associated with rhythm perception, but not with melody perception. *Journal of Experimental Psychology: Human Perception and Performance*, *41*(2), 277.

Bidelman, G. M., Hutka, S., & Moreno, S. (2013). Tone language speakers and musicians share enhanced perceptual and cognitive abilities for musical pitch: Evidence for bidirectionality between the domains of language and music. *PloS One*, *8*(4).

Bigand, E., Parncutt, R., & Lerdahl, F. (1996). Perception of musical tension in short chord sequences: The influence of harmonic function, sensory dissonance, horizontal motion, and musical training. *Perception & Psychophysics*, *58*(1), 125–141.

Bochynska, A., Vulchanova, M., Vulchanov, V., & Landau, B. (2020). Spatial language difficulties reflect the structure of intact spatial representation: Evidence from high-functioning autism. *Cognitive Psychology*, *116*, 101249.

Burger, B., Thompson, M. R., Luck, G., Saarikallio, S. H., & Toiviainen, P. (2014). Hunting for the beat in the body: On period and phase locking in music-induced movement. *Frontiers in Human Neuroscience*, *8*, 903.

Bybee, J. (2006). From usage to grammar: The mind’s response to repetition. *Language*, *82*(4), 711–733. Retrieved from <http://www.jstor.org/stable/4490266>

Carey, D., Rosen, S., Krishnan, S., Pearce, M. T., Shepherd, A., Aydelott, J., & Dick, F. (2015). Generality and specificity in the effects of musical expertise on perception and cognition. *Cognition*, *137*, 81–105.

Carpenter, P. A., & Just, M. A. (1977). Reading comprehension as eyes see it. *Cognitive Processes in Comprehension*, 109–139.

Cason, N., Marmursztejn, M., D’Imperio, M., & Schön, D. (2019). Rhythmic abilities correlate with l2 prosody imitation abilities in typologically different languages. *Language and Speech*, 0023830919826334.

Cason, N., & Schön, D. (2012). Rhythmic priming enhances the phonological processing of speech. *Neuropsychologia*, *50*(11), 2652–2658.

Chan, R. K., & Leung, J. H. (2019). WHY are lexical tones difficult to learn?: INSIGHTS from the incidental learning of tone-segment connections. *Studies in Second Language Acquisition*, 1–27.

Chao, Y. R. (1968). A grammar of spoken chinese.

Chen, H. C. (2013). Chinese learners’ acquisition of english word stress and factors affecting stress assignment. *Linguistics and Education*, *24*(4), 545–555.

Child, S., Oakhill, J., & Garnham, A. (2020). Tracking your emotions: An eye-tracking study on reader’s engagement with perspective during text comprehension. *Quarterly Journal of Experimental Psychology*, 1747021820905561.

Chomsky, N. (1986). *Knowledge of language: Its nature, origin, and use*. Greenwood Publishing Group.

Christiansen, M. H., & Chater, N. (2016). The now-or-never bottleneck: A fundamental constraint on language. *Behavioral and Brain Sciences*, *39*.

Chua, A. J., & Brunt, J. (2014). Effect of musical experience on tonal language perception. In *Proceedings of meetings on acoustics 167ASA* (Vol. 21, p. 060009). Acoustical Society of America.

Colantoni, L., Steele, J., Escudero, P., & Neyra, P. R. E. (2015). *Second language speech*. Cambridge University Press.

Colley, I. D., Keller, P. E., & Halpern, A. R. (2018). Working memory and auditory imagery predict sensorimotor synchronisation with expressively timed music. *Quarterly Journal of Experimental Psychology*, *71*(8), 1781–1796.

Cooper, N., Cutler, A., & Wales, R. (2002). Constraints of lexical stress on lexical access in english: Evidence from native and non-native listeners. *Language and Speech*, *45*(3), 207–228.

Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology*.

Correia, S., Frota, S., Butler, J., & Vigário, M. (2013). Word stress perception in european portuguese. In *INTERSPEECH* (pp. 267–271).

Cowan, N. (2016). *Working memory capacity: Classic edition*. Psychology press.

Cuetos, F., Mitchell, D. C., & Corley, M. M. (1996). Language processing in spanish. In (pp. 145–187). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Cutler, A. (1986). Forbear is a homophone: Lexical prosody does not constrain lexical access. *Language and Speech*, *29*(3), 201–220.

Deutsch, D., Dooley, K., Henthorn, T., & Head, B. (2009). Absolute pitch among students in an american music conservatory: Association with tone language fluency. *The Journal of the Acoustical Society of America*, *125*(4), 2398–2403.

Dijkgraaf, A., Hartsuiker, R. J., & Duyck, W. (2017). Predicting upcoming information in native-language and non-native-language auditory word recognition. *Bilingualism: Language and Cognition*, *20*(5), 917–930.

Duchowski, A. T. (2007). Eye tracking methodology. *Theory and Practice*, *328*(614), 2–3.

Dupoux, E., Sebastián-Gallés, N., Navarrete, E., & Peperkamp, S. (2008). Persistent stress “deafness”: The case of french learners of spanish. *Cognition*, *106*(2), 682–706.

Dussias, P. E., Kroff, J. R. V., Tamargo, R. E. G., & Gerfen, C. (2013). When gender and looking go hand in hand: Grammatical gender processing in l2 spanish. *Studies in Second Language Acquisition*, *35*(2), 353–387.

Face, T. L. (2006). Cognitive factors in the perception of spanish stress placement: Implications for a model of speech perception. *Linguistics*, *44*(6), 1237–1267.

Federmeier, K. D. (2007). Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology*, *44*(4), 491–505.

Fischer, B. (1992). Saccadic reaction time: Implications for reading, dyslexia, and visual cognition. In *Eye movements and visual cognition* (pp. 31–45). Springer.

Fodor, J. A. (1983). *The modularity of mind*. MIT press.

Fraisse, P. (2013). Psychology of music. In D. Deutsch (Ed.). Elsevier.

Fujioka, T., Ross, B., & Trainor, L. J. (2015). Beta-band oscillations represent auditory beat and its metrical hierarchy in perception and imagery. *Journal of Neuroscience*, *35*(45), 15187–15198.

Gandour, J. T., & Fromkin, V. (1978). The perception of tone in tone: A linguistic survey.

Ghaffarvand Mokari, P., & Werner, S. (2018). Perceptual training of second-language vowels: Does musical ability play a role? *Journal of Psycholinguistic Research*, *47*(1), 95–112.

Grabe, E., & Low, E. L. (2002). Durational variability in speech and the rhythm class hypothesis. *Papers in Laboratory Phonology*, *7*(515-546).

Grüter, T., & Rohde, H. (2013). L2 processing is affected by rage: Evidence from reference resolution. In *The 12th conference on generative approaches to second language acquisition (gasla)*.

Hasson, U., Egidi, G., Marelli, M., & Willems, R. M. (2018). Grounding the neurobiology of language in first principles: The necessity of non-language-centric explanations for language comprehension. *Cognition*, *180*, 135–157.

Hawthorne, K., & Gerken, L. (2014). From pauses to clauses: Prosody facilitates learning of syntactic constituency. *Cognition*, *133*(2), 420–428.

Heard, M., & Lee, Y. S. (2020). Shared neural resources of rhythm and syntax: An ale meta-analysis. *Neuropsychologia*, *137*, 107284.

Hed, A., Schremm, A., Horne, M., & Roll, M. (2019). Neural correlates of second language acquisition of tone-grammar associations. *The Mental Lexicon*, *14*(1), 98–123.

Hopp, H. (2013). Grammatical gender in adult l2 acquisition: Relations between lexical and syntactic variability. *Second Language Research*, *29*(1), 33–56.

Hopp, H. (2015). Semantics and morphosyntax in predictive l2 sentence processing. *International Review of Applied Linguistics in Language Teaching*, *53*(3), 277–306.

Hopp, H. (2016). Learning (not) to predict: Grammatical gender processing in second language acquisition. *Second Language Research*, *32*(2), 277–307.

Hothorn, T., Bretz, F., Westfall, P., Heiberger, R. M., Schuetzenmeister, A., Scheibe, S., & Hothorn, M. T. (2016). Package “multcomp”. *Simultaneous Inference in General Parametric Models. Project for Statistical Computing, Vienna, Austria*.

Hu, J., Gao, S., Ma, W., & Yao, D. (2012). Dissociation of tone and vowel processing in m andarin idioms. *Psychophysiology*, *49*(9), 1179–1190.

Hualde, J. I. (2005). *The sounds of spanish with audio cd*. Cambridge University Press.

Hualde, J. I., & Kim, J. Y. (2015). The acquisition of spanish lexical stress by korean learners. In *ICPhS*.

Huettig, F., & Mani, N. (2016). Is prediction necessary to understand language? Probably not. *Language, Cognition and Neuroscience*, *31*(1), 19–31.

Huettig, F., & McQueen, J. M. (2007). The tug of war between phonological, semantic and shape information in language-mediated visual search. *Journal of Memory and Language*, *57*(4), 460–482.

Huettig, F., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica*, *137*(2), 151–171.

Ito, A. (2016). Prediction during native and non-native language comprehension: The role of mediating factors.

Kamide, Y., Altmann, G. T., & Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language*, *49*(1), 133–156.

Kamin, L. (1969). Predictability, surprise, attention and cognition. *BA Campbell, RM Church New York: Appleton-Century-Crofts*.

Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working memory capacity: A latent-variable approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology: General*, *133*(2), 189.

Kjelgaard, M. M., & Speer, S. R. (1999). Prosodic facilitation and interference in the resolution of temporary syntactic closure ambiguity. *Journal of Memory and Language*, *40*(2), 153–194.

Klapuri, A. (2006). Introduction to music transcription. In *Signal processing methods for music transcription* (pp. 3–20). Springer.

Lee, G., Shin, D.-J., & Garcia, M. T. M. (2019). Perception of lexical stress and sentence focus by korean-speaking and spanish-speaking l2 learners of english. *Language Sciences*, *72*, 36–49.

Lew-Williams, C., & Fernald, A. (2010). Real-time processing of gender-marked articles by native and non-native spanish speakers. *Journal of Memory and Language*, *63*(4), 447–464.

Li, X., To, C. K. S., & Ng, M. L. (2017). Effects of l1 tone on perception of l2 tone-a study of mandarin tone learning by native cantonese children. *Bilingualism: Language and Cognition*, *20*(3), 549–560.

Liburd, T. (2014). *Investigating prediction in l2 morpho-syntax: A visual world study* (PhD thesis). University of Pittsburgh.

Lin, H. 林., & Wang, Q. 王倩. (2007). Mandarin rhythm: An acoustic study. *Journal of Chinese Language and Computing*, *17*(3), 127–140.

Linck, J. A., Osthus, P., Koeth, J. T., & Bunting, M. F. (2014). Working memory and second language comprehension and production: A meta-analysis. *Psychon. Bull. Rev.*, *21*(4), 861–883.

Lord, G. (2007). The role of the lexicon in learning second language stress patterns. *Applied Language Learning*, *17*(1/2), 1.

Loui, P., & Wessel, D. (2007). Harmonic expectation and affect in western music: Effects of attention and training. *Perception & Psychophysics*, *69*(7), 1084–1092.

Lozano-Argüelles, C. (2020). *Predicting the end: Monolilnguals, l2 learners and interpreters’ use of prosody to predict word endings*.

Lozano-Argüelles, C., Sagarra, N., & Casillas, J. V. (2019). Slowly but surely: Interpreting facilitates l2 morphological anticipation based on suprasegmental and segmental information. *Bilingualism: Language and Cognition*, 1–11.

Ma, S. M. (2016). Working memory in spanish-english and chinese-english bilinguals. *Psychology and Behavioral Sciences*, *5*(4), 104–112.

Mackintosh, N. J. (1975). A theory of attention: Variations in the associability of stimuli with reinforcement. *Psychological Review*, *82*(4), 276.

Magne, C., Jordan, D. K., & Gordon, R. L. (2016). Speech rhythm sensitivity and musical aptitude: ERPs and individual differences. *Brain and Language*, *153*, 13–19.

Magne, C., Schön, D., & Besson, M. (2006). Musician children detect pitch violations in both music and language better than nonmusician children: Behavioral and electrophysiological approaches. *Journal of Cognitive Neuroscience*, *18*(2), 199–211.

Malins, J. G., & Joanisse, M. F. (2010). The roles of tonal and segmental information in mandarin spoken word recognition: An eyetracking study. *Journal of Memory and Language*, *62*(4), 407–420.

Marques, C., Moreno, S., Luı's Castro, S., & Besson, M. (2007). Musicians detect pitch violation in a foreign language better than nonmusicians: Behavioral and electrophysiological evidence. *Journal of Cognitive Neuroscience*, *19*(9), 1453–1463.

Martin, J. G. (1972). Rhythmic (hierarchical) versus serial structure in speech and other behavior.

Marull, C. (2017). Second language processing efficiency: Experience and cognitive effects on l2 morphosyntactic integration and anticipation. In *BUCLD 41: Proceedings of the 41st boston university conference on language development* (pp. 466–480). Cascadilla Press Somerville, MA.

McDermott, J. H., Lehr, A. J., & Oxenham, A. J. (2010). Individual differences reveal the basis of consonance. *Current Biology*, *20*(11), 1035–1041.

Milner, B. (1971). Interhemispheric differences in the localization of psychological processes in man. *British Medical Bulletin*.

Mirman, D. (2016). *Growth curve analysis and visualization using r*. CRC press.

Mitsugi, S., & Macwhinney, B. (2016). The use of case marking for predictive processing in second language japanese. *Bilingualism: Language and Cognition*, *19*(1), 19–35.

Miyake, Y., Onishi, Y., & Poppel, E. (2004). Two types of anticipation in synchronization tapping. *Acta Neurobiologiae Experimentalis*, *64*(3), 415–426.

Mok, P. (2009). On the syllable-timing of cantonese and beijing mandarin. *Chinese Journal of Phonetics*, *2*, 148–154.

Morando, A., Victor, T., & Dozza, M. (2016). Drivers anticipate lead-vehicle conflicts during automated longitudinal control: Sensory cues capture driver attention and promote appropriate and timely responses. *Accident Analysis & Prevention*, *97*, 206–219.

Morrison, S. J., & Demorest, S. M. (2009). Cultural constraints on music perception and cognition. *Progress in Brain Research*, *178*, 67–77.

Musacchia, G., Sams, M., Skoe, E., & Kraus, N. (2007). Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. *Proceedings of the National Academy of Sciences*, *104*(40), 15894–15898.

Nakamoto, H., & Mori, S. (2012). Experts in fast-ball sports reduce anticipation timing cost by developing inhibitory control. *Brain and Cognition*, *80*(1), 23–32.

Nakamura, C., Arai, M., & Mazuka, R. (2012). Immediate use of prosody and context in predicting a syntactic structure. *Cognition*, *125*(2), 317–323.

Ngo, M. K., Vu, K.-P. L., & Strybel, T. Z. (2016). Effects of music and tonal language experience on relative pitch performance. *The American Journal of Psychology*, *129*(2), 125–134.

Ortega-Llebaria, M. (2006). Phonetic cues to stress and accent in spanish. In *Selected proceedings of the 2nd conference on laboratory approaches to spanish phonetics and phonology* (pp. 104–118). Cascadilla Proceedings Project Somerville, MA.

Ortega-Llebaria, M., Gu, H., & Fan, J. (2013). English speakers’ perception of spanish lexical stress: Context-driven l2 stress perception. *Journal of Phonetics*, *41*(3-4), 186–197.

Ortega-Llebaria, M., & Prieto, P. (2007). Disentangling stress from accent in spanish: Production patterns of the stress contrast in deaccented syllables. *Amsterdam Studies in the Theory and History of Linguistic Science Series 4*, *282*, 155.

Ortega-Llebaria, M., & Prieto, P. (2009). Perception of word stress in castilian spanish. *Phonetics and Phonology: Interactions and Interrelations*, *306*, 35.

Otten, M., & Van Berkum, J. J. (2009). Does working memory capacity affect the ability to predict upcoming words in discourse? *Brain Research*, *1291*, 92–101.

Pagliarini, E. (2016). Predictive timing in developmental dyslexia: A new hypothesis. Anticipatory skills across language and motor domains.

Pakulak, E., & Neville, H. J. (2010). Proficiency differences in syntactic processing of monolingual native speakers indexed by event-related potentials. *Journal of Cognitive Neuroscience*, *22*(12), 2728–2744.

Patel, A. D., & Iversen, J. R. (2014). The evolutionary neuroscience of musical beat perception: The action simulation for auditory prediction (asap) hypothesis. *Frontiers in Systems Neuroscience*, *8*, 57.

Pearce, M. T., & Wiggins, G. A. (2012). Auditory expectation: The information dynamics of music perception and cognition. *Topics in Cognitive Science*, *4*(4), 625–652.

Perdomo, M., & Kaan, E. (2019). Prosodic cues in second-language speech processing: A visual world eye-tracking study. *Second Language Research*, 0267658319879196.

Persici, V., Stucchi, N., & Arosio, F. (2019). Rhythmic and morphosyntactic predictions: The anticipation abilities of italian children with developmental dyslexia.

Phillips, C. E., Jarrold, C., Baddeley, A. D., Grant, J., & Karmiloff-Smith, A. (2004). Comprehension of spatial language terms in williams syndrome: Evidence for an interaction between domains of strength and weakness. *Cortex*, *40*(1), 85–101.

Phillips-Silver, J., Toiviainen, P., Gosselin, N., & Peretz, I. (2013). Amusic does not mean unmusical: Beat perception and synchronization ability despite pitch deafness. *Cognitive Neuropsychology*, *30*(5), 311–331.

Pozzan, L., Gleitman, L. R., & Trueswell, J. C. (2016). Semantic ambiguity and syntactic bootstrapping: The case of conjoined-subject intransitive sentences. *Language Learning and Development*, *12*(1), 14–41.

R Core Team. (2019). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org/>

Rehrig, G. L. (2017). *Acoustic correlates of syntax in sentence production and comprehension* (PhD thesis). Rutgers University-School of Graduate Studies.

Roden, I., Grube, D., Bongard, S., & Kreutz, G. (2014). Does music training enhance working memory performance? Findings from a quasi-experimental longitudinal study. *Psychology of Music*, *42*(2), 284–298.

Rohrmeier, M. A., & Koelsch, S. (2012). Predictive information processing in music cognition. A critical review. *International Journal of Psychophysiology*, *83*(2), 164–175.

Roll, M., & Horne, M. (2015). A neurolinguistic study of south swedish word accents: Electrical brain potentials in nouns and verbs, *38*(2), 149–162.

Roll, M., Horne, M., & Lindgren, M. (2010). Word accents and morphology—erps of swedish word processing. *Brain Research*, *1330*, 114–123.

Roll, M., Horne, M., & Lindgren, M. (2011). Activating without inhibiting: Left-edge boundary tones and syntactic processing. *Journal of Cognitive Neuroscience*, *23*(5), 1170–1179.

Roll, M., Söderström, P., & Horne, M. (2013). Word-stem tones cue suffixes in the brain. *Brain Research*, *1520*, 116–120.

Ryskin, R., Levy, R. P., & Fedorenko, E. (2020). Do domain-general executive resources play a role in linguistic prediction? Re-evaluation of the evidence and a path forward. *Neuropsychologia*, *136*, 107258.

Sagarra, N., & Casillas, J. V. (2018). Suprasegmental information cues morphological anticipation during l1/l2 lexical access. *Journal of Second Language Studies*, *1*(1), 31–59.

Sagarra, N., & Herschensohn, J. (2010). The role of proficiency and working memory in gender and number agreement processing in l1 and l2 spanish. *Lingua*, *120*(8), 2022–2039.

Salverda, A. P., Kleinschmidt, D., & Tanenhaus, M. K. (2014). Immediate effects of anticipatory coarticulation in spoken-word recognition. *Journal of Memory and Language*, *71*(1), 145–163.

Saslow, M. (1967). Effects of components of displacement-step stimuli upon latency for saccadic eye movement. *Josa*, *57*(8), 1024–1029.

Schellenberg, E. G. (1996). Expectancy in melody: Tests of the implication-realization model. *Cognition*, *58*(1), 75–125.

Schellenberg, E. G., & Habashi, P. (2015). Remembering the melody and timbre, forgetting the key and tempo. *Memory & Cognition*, *43*(7), 1021–1031.

Schön, D., Gordon, R. L., & Besson, M. (2005). Musical and linguistic processing in song perception. *Annals of the New York Academy of Sciences*, *1060*(1), 71–81.

Schremm, A., Söderström, P., Horne, M., & Roll, M. (2016). Implicit acquisition of tone-suffix connections in l2 learners of swedish. *The Mental Lexicon*, *11*(1), 55–75.

Schwab, S., & Dellwo, V. (2016). The use of the odd-one-out task in the study of the perception of lexical stress in spanish by german-speaking listeners?

Shen, W., Hyönä, J., Wang, Y., Hou, M., & Zhao, J. (2020). The role of tonal information during spoken-word recognition in chinese: Evidence from a printed-word eye-tracking study. *Memory & Cognition*, 1–12.

Slevc, L. R., Davey, N. S., Buschkuehl, M., & Jaeggi, S. M. (2016). Tuning the mind: Exploring the connections between musical ability and executive functions. *Cognition*, *152*, 199–211.

Slevc, L. R., & Miyake, A. (2006). Individual differences in second-language proficiency: Does musical ability matter? *Psychological Science*, *17*(8), 675–681.

Soto-Faraco, S., Sebastián-Gallés, N., & Cutler, A. (2001). Segmental and suprasegmental cues for lexical access in spanish. *Journal of Memory and Language*, *45*, 412–432.

Söderström, P., Horne, M., & Roll, M. (2015). Using tonal cues to predict inflections. *Working Papers 55. 2015 Linguistics Lund University*, 91.

Söderström, P., Roll, M., & Horne, M. (2012). Processing morphologically conditioned word accents. *The Mental Lexicon*, *7*(1), 77–89.

Stahl, P., Donmez, B., & Jamieson, G. A. (2016). Supporting anticipation in driving through attentional and interpretational in-vehicle displays. *Accident Analysis & Prevention*, *91*, 103–113.

Staub, A., & Clifton Jr, C. (2006). Syntactic prediction in language comprehension: Evidence from either... Or. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(2), 425.

Strait, D. L., Kraus, N., Parbery-Clark, A., & Ashley, R. (2010). Musical experience shapes top-down auditory mechanisms: Evidence from masking and auditory attention performance. *Hearing Research*, *261*(1-2), 22–29.

System, V. T. (2013). Zeit- und bewegungsantizipation “time and movement anticipation”.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*(5217), 1632–1634.

Tang, W., Xiong, W., Zhang, Y.-x., Dong, Q., & Nan, Y. (2016). Musical experience facilitates lexical tone processing among mandarin speakers: Behavioral and neural evidence. *Neuropsychologia*, *91*, 247–253.

Tierney, A., & Kraus, N. (2013). The ability to move to a beat is linked to the consistency of neural responses to sound. *Journal of Neuroscience*, *33*(38), 14981–14988.

Tillmann, B. (2012). Music and language perception: Expectations, structural integration, and cognitive sequencing. *Topics in Cognitive Science*, *4*(4), 568–584.

Tsukada, K., Xu, H. L., & Rattanasone, N. X. (2015). The perception of mandarin lexical tones by listeners from different linguistic backgrounds. *Chinese as a Second Language Research*, *4*(2), 141–161.

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, *37*(3), 498–505.

Van de Cavey, J., & Hartsuiker, R. J. (2016). Is there a domain-general cognitive structuring system? Evidence from structural priming across music, math, action descriptions, and language. *Cognition*, *146*, 172–184.

Vidyasagar, T. R., & Pammer, K. (2010). Dyslexia: A deficit in visuo-spatial attention, not in phonological processing. *Trends in Cognitive Sciences*, *14*(2), 57–63.

Weber, A., Grice, M., & Crocker, M. W. (2006). The role of prosody in the interpretation of structural ambiguities: A study of anticipatory eye movements. *Cognition*, *99*(2), B63–B72.

Wicha, N. Y., Bates, E. A., Moreno, E. M., & Kutas, M. (2003). Potato not pope: Human brain potentials to gender expectation and agreement in spanish spoken sentences. *Neuroscience Letters*, *346*(3), 165–168.

Wiener, S., & Turnbull, R. (2016). Constraints of tones, vowels and consonants on lexical selection in mandarin chinese. *Language and Speech*, *59*(1), 59–82.

Wilson, M., & Lindstrom, S. H. (2011). What the bird’s brain tells the bird’s eye: The function of descending input to the avian retina. *Visual Neuroscience*, *28*(4), 337–350.

Wong, P. C., Skoe, E., Russo, N. M., Dees, T., & Kraus, N. (2007). Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nature Neuroscience*, *10*(4), 420–422.

Wright, M. J., Bishop, D. T., Jackson, R. C., & Abernethy, B. (2010). Functional mri reveals expert-novice differences during sport-related anticipation. *Neuroreport*, *21*(2), 94–98.

Yang, E. (2017). Bilinguals’ working memory (wm) advantage and their dual language practices. *Brain Sciences*, *7*(7), 86.

Yu, M., Xu, M., Li, X., Chen, Z., Song, Y., & Liu, J. (2017). The shared neural basis of music and language. *Neuroscience*, *357*, 208–219.

Zhu, X., & Wang, C. (2015). The oxford handbook of chinese linguistics. In W. S. Wang & C. Sun (Eds.). Oxford University Press.