WHAT’S NEXT: THE ROLE OF RHYTHMIC, MELODIC AND VISUAL-SPATIAL ANTICIPATION ABILITIES, L1 TRANSFER AND L2 PROFICIENCY ON L2 ANTICIPATION OF VERB SUFFIXES

Laura Fernandez Arroyo1

1 Rutgers University

WHAT’S NEXT: THE ROLE OF RHYTHMIC, MELODIC AND VISUAL-SPATIAL ANTICIPATION ABILITIES, L1 TRANSFER AND L2 PROFICIENCY ON L2 ANTICIPATION OF VERB SUFFIXES

## Acoustic correlates

*Lexical stress*

Stress is the prominence of a syllable that speakers hear relative to the other syllables in the prosodic word (Hualde, 2005). Lexical stress has no fixed position in either English or Spanish and thus is phonologically contrastive at the lexical level in both languages. That is, lexical stress can be used to distinguish between words. The contrastive use is nevertheless much more typical in Spanish than in English. In English it is used predominantly to distinguish English heteronyms or pairs of verbs-nouns that have no segmental differences. For example, to “PROduce,” verb vs. “proDUCE,” noun. 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”). The major emphasis of the stressed syllable is caused by different acoustic correlates depending on the language. Stress is the combined result of many parameters in action, among which we can find F0 variations, duration, overall intensity, and vowel formant frequencies (Gordon & Roettger, 2017), and the different importances or weights assigned to each of these correlates cause the nature of stressed syllables in each language to vary. In Spanish, the most reliable cues to stress are pitch (F0) height, duration and intensity (Hualde, 2005; Ortega-Llebaria, 2006; Ortega-Llebaria & Prieto, 2007, 2009). Pitch is high for stressed syllables and low 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), although other cues such as intensity and pitch (F0] height (Beckman, 1986; Fry, 1955, 1958, 1965; Sluijter & Van Heuven, 1996; Sluijter, Van Heuven, & Pacilly, 1997) play minor roles. Thus, unstressed vowels are reduced to [ə] and their formants are centralized. The different weight assigned to each cue in these languages may explain why L1 English speakers encounter difficulties in Spanish lexical stress perception (Face, 2006; Ortega-Llebaria, Gu, & Fan, 2013) and production (Lord, 2007). The different cue weights have also effects on the prosodic structure of the language. Vowel reduction causes English to be a stressed-timed language, where the intervals between stressed syllables have similar durations. Spanish, on the contrary, is a syllable-timed language, as all syllables last approximately the same 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; while in Spanish, the stress is simply part of a syllable, and each syllable is a new rhythmic unit. The different weights assigned to each cue in different languages may also be one of the factors determining the differences in lexical stress processing particular to each language. As mentioned above, lexical stress helps activation of lexical entries in L1 Spanish (Soto-Faraco, Sebastián-Gallés, & Cutler, 2001), such that a prosodically matching cue to the target (prinCI > prinCIpio, “start”) results in shorter and more accurate decision making times, when compared to mismatching cues (PRINci > prinCIpio). These results are taken to suggest that participants in Soto-Faraco et al. (2001) study were anticipating the lexical element based on suprasegmental cues such as lexical stress. Contrarily, it is unclear whether L1 English speakers are able to use stress alone as a cue to anticipate and facilitate lexical activation. On the one hand, Cooper et al. (2002) tested L1 English speakers in a similar study to that of Soto-Faraco et al. (2001) and found that the English natives were only able to use the suprasegmental cues when more than one syllable of the word was present. On the other hand, Perdomo and Kaan (2019) did find in an eye-tracking study that the presence of lexical stress elicited fixations on the oncoming target noun. It is possible that the L1 English speakers were using low emphasis in previous syllables to activate the cue role for lexical stress; similarly to what the L1 Spanish might have done, as the target words came at the end of a context sentence. This difference in performance amongst studies is probably due to the weights assigned to each cue and how it is presented. That is, to the larger reliance of English L1 speakers on duration to process Spanish lexical stress transferred from L1 English processing against the reliance of Spanish L1 speakers on other cues, such as pitch and intensity, that are discarded by the English speakers (Ortega-Llebaria et al. (2013)). The differences and similarities in cue weighting can no doubt 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. But German speakers perceive stress significantly more poorly in another free lexically stressed language such as Spanish than L1 speakers (Schwab & Dellwo, 2016); again, maybe due to the different weights assigned to each cue. Following this line, Vickie and Andruski (2010) suggested that language background in the L1 can affect how lexical stress is perceived in an L2, when their L1 Mandarin participants used different cues to process L2 English lexical stress as compared to L1 English participants. The influence of the different weights assigned to each cue might extend beyond perception and affects anticipation as well. The studies above suggest that the acoustic properties of prosody are essential in processing and activating language. Studies like the one by Vickie and Andruski (2010) further suggest that speakers can resort to their prosodic abilities in the L1 in order to process other prosodic structures in their L2 absent in their L1. That would be the case of L1 speakers of tonal languages learning L2 stressed languages. Extending this hypothesis, L2 speakers might be able to reassemble prosodic structure and its basis in order to promote L2 anticipation so as to reduce the processing cognitive load.

*Tones*

Tones are the pitch contour patterns of the voiced part in syllables (Chao, 1968). Many languages use tones, or changes in pitch-contour, at a phrasal level for pragmatic purposes. However, only a few use tones contrastively at a lexical level. The acoustic correlates for tones vary across languages: some use only pitch (e.g., Mandarin Chinese), whereas others also use length and/or register (e.g., Cantonese Chinese). Relevant to my dissertation with Mandarin Chinese speakers, in most Mandarin Chinese dialects (e.g., from Beijing and Tianjin), the main acoustic correlate for tones is changes in pitch (F0) contour or changes in pitch height within a syllable (Gandour & FROMKIN, 1978; Zhu & Wang, 2015). Importantly, tones in Mandarin Chinese do not cause shorter and longer syllables. Therefore, Mandarin is 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 (Malins and Joanisse (2010)). They are nevertheless not the most important factor in that process, as vowel especially but also consonant identity comes first (Hu, Gao, Ma, & Yao, 2012). In other words, while tones confirm that the correct word is activated, the main vehicle to access a lexicon entry in Mandarin are other cues, mainly segmental cues. Wiener and Turnbull (2016) investigated the degree to which segmental (vocalic and consonantal phonemes) and suprasegmental cues (tones) constrained lexical access in Mandarin Chinese. L1 Mandarin speakers were presented different types of stimuli containing different types of violations (tonotactic, phonactic) and had to decide as fast as possible if what they were hearing was a real word or not. Words with tonotactic violations were more often endorsed as real words than other types of violations, such as vowel or consonant change. These results led the authors to conclude that tone information is not as important as consonant and especially vowel information in lexical access. Hu et al. (2012) reached an akin conclusion in a ERP study. Hu et al. (2012) studied the relevance of tone and vowel information at different stages of lexical access, for which they selected fixed Chinese idioms and isolated words as context. Vowel mismatches evoked earlier (N1 effect) than tone mismatches (N400). The N1 effect was taken to suggest that vowel was playing a role on word selection, while the N400 effect signals a failure of the integration of the word in the sentential context. Sereno and Lee (2015) obtained similar results in two priming experiments in Mandarin. Participants were slower in discriminating words where the only difference was tonal. Primings where both vowel and tone matched where the fastest one, followed by primings where only vowel matched. The three studies led to the conclusion that the functions of vowels and tones in Mandarin are distinct. Namely, that vowel plays a major role in activation, while the role of tone is integration in the higher context. The knowledge of the nature and function of tones in the L1 can affect L2 tone learning positively by providing a background knowledge to which learners can resort to acquire the L2 tones. However, L1 tone knowledge can also affect L2 tone learning negatively when transferring the L1 tone knowledge interferes with the nature of the L2 tones. Li, To, and Ng (2017) examined the influence of the L1 tonal knowledge in the acquisition of L2 tones in children. These children were L1 Cantonese speakers learning L2 Mandarin, and they had issues in categorizing Mandarin tones 1 and 4, as these tones would be assimilated to the same tone 1 category in Cantonese. In the case of these children, being a native speaker to a tonal language helped them in the perception of Mandarin tones 2 and 3, but it hindered perception of other L2 tones because the knowledge transfer from L1 tones disagreed from the L2 tone structure and interfered with it. Although it seems the role of tones is not as pre-eminent of lexical stress in Spanish, it still helps in word activation and must be encoded in the lexicon. Mandarin speakers need to pay attention to the pitch variations in order to assign the correct tone to the word they are hearing. Since pitch variations are the basis for lexical stress in Spanish, Mandarin speakers might be able to transfer their sensitivity to pitch changes to process and use pitch to anticipate linguistic information more easily than English speakers. Although English also has lexical stress, it discards the acoustic correlate, so learning to distinguish the pitch variations may be more difficult than simply transferring the sensitivity, and thus Mandarin speakers may outperform English speakers in using lexical stress to anticipate verbal tense in L2 Spanish.

# Music

There is abundant evidence supporting a cognitive association between music and language. From a visual standpoint, higher musical abilities yield more effective reading comprehension skills in children (Anvari, Trainor, Woodside, & Levy, 2002). Regarding acoustic processing, musical abilities influence positively how linguistic sound is encoded in adults (Wong, Skoe, Russo, Dees, & Kraus, 2007) as well as in children (Magne, Schön, & Besson, 2006), and also in a foreign language (Marques, Moreno, Luı's Castro, & Besson, 2007; Slevc & Miyake, 2006). The positive effect of musical experience on language is also reflected on enhanced cortical ERPs: Even when shared acoustic resources, like pitch, are kept stable across conditions, musicians process faster speech than non-musicians and the brainstem responses happen earlier (Musacchia, Sams, Skoe, & Kraus, 2007). The influence of music in language starts early during childhood. Regarding syntax, Jentschke, Friederici, and Koelsch (2014) found that music and language follow a similar time course of acquisition of syntactic regularities. Degé and Schwarzer (2011) compared the effects of a music program and a phonological skills program on children’s phonological awareness, and they found that phonological awareness of small and large phonological units improved from a pre-test to a post-test for both treatment groups, while no improvement was observed for a control group. In a similar fashion, Lebedeva and Kuhl (2010) explored whether infants were able to discriminate phonetic from melodic patterns in songs. Results indicated that 11-month-old infants could discriminate between 4-note sung melodies, but the infants did not show any preference towards familiar or unfamiliar spoken strings containing a syllable order change when the intonation contour remained stable for both strings. However, when the melodies were presented without the phonetic cue, and thus were not sung only instrumental, infants did not show any preference, suggesting that they were not discriminating the melodies based on their melodic contour alone. Taken together, the findings of the tasks Lebedeva and Kuhl (2010) deployed suggest that infants use the phonetic content of melodies to process a melodic line, even if they are not able to understand fully what they are hearing. Findings like the one by Lebedeva and Kuhl (2010) and the resemblance in processing between music and language have even led some scholars to suggest that language is processed as a type of music by the early developing brain (Koelsch & Siebel, 2005). The music-language association found in children is also reported in adult native speakers and L2 learners. ERPs have shown that music-syntactic and language-syntactic irregularities produce similar reduced responses in amusic individuals, while semantic irregularities were unaffected (Sun et al., 2018). 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 cancelation test and an onset cancelation test) and music tests (an interval test and a rhythmic test). The training group performed better in all tests, and what is more, the accuracy in the animal-word cancelation test and the accuracy in the interval test were positively correlated. With regard to L2 learners, Ghaffarvand Mokari and Werner (2018) studied whether there was any kind of association between general musical abilities and L2 vowel learning with L1 Azerbaijani learners of English. The participants were split into control group and experimental group, who underwent phonetic training. Both groups completed music ability tests (chord analysis, pitch change and tonal memory) and linguistic tests (discrimination and production). Results point that the increased accuracy in discrimination and production of L2 sounds from the pre- to the post-test is not related to the musical ability in general, although tonal memory was significantly correlated to the gained skill to discriminate L2 vowels. Regarding rhythm, Cason, Marmursztejn, D’Imperio, and Schön (2019) investigated how rhythmic abilities were related to the perception and production of L2 stress placement. Rhythmic production scores were a reliable predictor of L2 stress placement. Swaminathan, Schellenberg, and Venkatesan (2018) explored if and how rhythm perception was associated with reading abilities in both the L1 and L2, and they found however that better rhythm perception was not associated with increased reading abilities in either type of language. In sum, research suggests that at least some cognitive faculties underlying music and language are shared. Sharing the faculty might be determined by the area we are focusing on, like speech vs reading. Previous research also suggests that anticipation is an important mechanism in both realms, especially in relation to rhythm, but to the best of my knowledge, no study so far has tested whether this processing mechanism is also interdependent to each domain or shared. Positive correlation between linguistic anticipation and other types of anticipation common in music, such as rhythmic or tonal, could indicate that there are domain-general predictive cognitive mechanisms underpinning auditory anticipation as a whole.

## Pitch

Pitch is the frequency associated to a sound wave; this frequency places the sound within a scale ranging from low to high in perception (Klapuri, 2006). Pitch in music plays a similar role to that it plays in language. In language, a listener must process the melodic contour of speech, that is, the changes up and down of pitch in order to understand the meaning conveyed by prosody (Dilley, 2005). For example, a rising pitch contour signals a question (“Coffee?”), while a falling intonation signals a statement (“Coffee.”). Apart from the phrasal and sentential intonation, pitch also affects tones and predominance against other syllables within a word. In music, pitch affects the note we hear. Both domains are related, and extensive research with atypical populations like amusics or individuals with Williams Syndrome (WS) have been one of the main foci proving that relationship. Research with typical population has also provided evidence that pitch processing in music is related to linguistic abilities, but not always. Starting with typical populations, L1 speakers of tonal languages are better and faster at discriminating and detecting absolute pitch in music (e.g., Chua and Brunt (2014); Deutsch, Dooley, Henthorn, and Head (2009); Tsukada, Xu, and Rattanasone (2015)), since tones in music depend on pitch, like tones in many languages. Although music in different cultures around the world vary on how their scales are organized (as well as on the patterns of strong-weak pulses, Morrison and Demorest (2009)), the basis for the scales is always pitch. Pitch collections may be different from one music culture to another, but the range is shared. Bidelman, Hutka, and Moreno (2013) tested how L1 Cantonese listeners, non-tonal speaking musically-trained participants and a non-tonal speaking non-musician control group performed in a three-alternative forced-choice task, an auditory inspection time paradigm, a short-term pitch memory task and a melody discrimination task to compare their measures of auditory pitch acuity, music perception, and general cognitive ability. The Cantonese speakers outperformed the control group on most pitch and music perception measures, and performed similarly to the musicians’ group. Because of the good results for the Cantonese group, who had no musical knowledge, the authors suggested that music and language abilities transfer bidirectionally. Additionally, musicianship contributes to better lexical tone perception even in a tonal language (Tang, Xiong, Zhang, Dong, & Nan, 2016). In contrast to the studies described so far, other studies have failed to obtain a music-language association in typical associations. Chan and Leung (2019) studied the creation of L2 tone-segment connections in musicians and non-musicians. Their results indicate that musical experience did not promote tone-segment connection in a tonal L2 when the L1 was non-tonal, but a tonal L1 did promote L2 tone-segment connection and the creation of a rule-like system of L2 tone information. Likewise, only musical experience can be beneficial in relative pitch perception in music, while tonal-language experience has no effect in relative pitch perception performance (Ngo, Vu, & Strybel, 2016). From this set of results, it is possible to deduce that there is a limit in the reciprocal influence between music and language (musical and linguistic pitch for the purpose of this project), but where this limit lies has not been ascertained. An area the relationship between pitch and language is clearer is atypical populations. Research on atypical populations like amusics or individuals with WS has produced larger evidence of a connection between pitch and language.In a study in amusia (defined as reduced or lost musical ability, such that comprehension and production of music, and ability to read and/or write musical notation are impeded, Pearce (2005)), Patel, Wong, Foxton, Lochy, and Peretz (2008) observed that amusic individuals distinguish changes in pitch but are unable to detect the direction of the change, such that they can discriminate changes in intonation, but cannot really tell what the prosodic change means. This ability to perceive pitch and its changes is probably one of the most studied aspects of the relationship between music and language. Further evidence of shared cognitive mechanisms between music and language is provided by research on atypical populations. Martı'nez-Castilla and Sotillo (2014) investigated whether children with WS processed pitch in music and language through common mechanisms. For that purpose, the WS children and typically developing children completed a musical pitch discrimination task, a short-item discrimination task and a long-item discrimination task. The WS group performed well above chance in all tasks, but their scores were significantly lower than those of the control group of typically developing children. In the case of the experimental group, the scores for the musical pitch discrimination-task and the short-item discrimination task were also correlated. In sum, some areas of music and language are connected, but influence from music to language and vice versa may be blocked. What are these specific areas is not completely known. While the shared processing mechanisms between music and language have been previously researched, no study so far has looked into how the anticipation mechanisms may be shared by the different auditory domains, or how the different acoustic correlates may affect anticipation in these domains. Separately, there is evidence for the existence of prediction processes in music (e.g., Salimpoor, Zald, Zatorre, Dagher, and McIntosh (2015)) as well as in language. Like in language, anticipation in music can be cued through syntactic structure (Sammler, Novembre, Koelsch, & Keller, 2013) or acoustic properties (Loui & Wessel, 2007). Anticipation has been particularly examined in rhythm, as synchronization to a rhythm is intrinsically keeping track of the interval structure, and therefore anticipation of the next event.

## Rhythm

Rhythm is defined as a pattern of recurrent time intervals that usually occur periodically (Berlyne, 1971). This period nature allows to predict the start of an interval or the next recurring event based on what has already been perceived (Fraisse, 2013; Martin, 1972). Rhythms can be perceived visually, as research with deaf individuals show. Iversen, Patel, Nicodemus, and Emmorey (2015) investigated how the accuracy of visual timing abilities and rhythmic perception are affected by developmental experience by comparing deaf and hearing adults in a tap synchronization task with three types of isochronous rhythmic stimuli: a flashing visual square, an animated bouncing ball and auditory tones. Results revealed that synchronization driven by a silent moving visual stimulus can be as accurate as that driven by sound, and that deaf individuals could synchronize their movements as well or even better than hearing individuals when presented with visual stimuli. However, rhythms are perceived primarily through our auditory pathways, like speech. Rhythms perceived auditorily are the ones that have been used to research hearing populations, regardless of whether the studies also included visual tasks, like reading. Swaminathan et al. (2018) studied the existence of an association between rhythm perception and reading ability in adults in their L1 and L2 and found no evidence. However, Cason and Schön (2012) did find that on-beat sequence primes yielded faster L1 speech processing than off-beat primes in adults through EEG recordings. The contradictory results obtained from these populations suggest that the effects of rhythm in hearing adults may not be as extensive as in deaf individuals in the visual domain. Maybe because hearing individuals do not need to hone their visual perception and processing abilities. Alternatively, there might be some “loss” in the transfer of rhythmic abilities from a visual domain to an auditory domain or vice versa, causing the relationship between rhythm and language in different modalities to be weaker. This conclusion would be supported by findings suggesting that within a modality, the relationship between rhythm and language, in this case speech, is stronger. Like in tone and pitch, there is evidence that rhythm is associated with language in children, L2 speakers, and atypical populations (beat-deafness, dyslexia and WS) within the auditory modality. Starting with children, Carr, White-Schwoch, Tierney, Strait, and Kraus (2014) had their typically developing children complete a series of linguistic, verbal memory, rhythmic and musical in general tasks to investigate the relationship between rhythmic synchronization and the encoding of syllable envelops in pre-schoolers. Those children who synchronized better also had better perceptual and cognitive language skills. They were faster at naming objects and color, and their neural encoding of the speech envelope was also better. Encoding of sound is also 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 speakers and L2 Spanish. 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. These studies suggest that rhythm is indeed related to speech in typical populations within a modality, and specifically, within the auditory domain. Studies with atypical populations have also yielded positive results supporting a relationship between heard rhythms and speech. Lagrois, Palmer, and Peretz (2019) tested beat-deaf individuals, that is, individuals who have a documented deficit in tracking musical beat, to analyze how they synchronized to speech. Participants had to tap along sentences in three conditions: regularly spoken, naturally spoken and sung, and were compared to a matched control group. The beat-deaf group showed more variability in its tapping regardless of the condition. The irregularity remained when they tapped to a metronome or at their own pace and was larger than in the typical non-musician control group. These results show that the deficiency in time-keeping mechanisms might underlie both domains rather than being domain-specific. In the same vein, Persici, Stucchi, and Arosio (2019) supported this conclusion when they compared the anticipation abilities of children with developmental dyslexia and typically developing children, and the results showed that children with dyslexia achieved lower scores than their peers in morphosyntactic and rhythmic processing. Therefore, noun prediction based on this information looked alike for both groups, but phonologically- and/or grammatically-based anticipation was hindered for the dyslexic children. Similarly, Sun et al. (2018) explored brain responses to syntactic violations in music and language in amusics. Brain responses to incongruities in both domains were reduced, but brain responses at later stages of processing, that is, during semantic processing, were unaffected or showed the same pattern as brain responses in typical individuals to semantic incongruencies. Based thus on previous findings suggesting the delayed anticipation in rhythm, phonology and morphosyntax, it is not unlikely that structure-based prediction is contingent to a cognitive mechanism common to the auditory mode in general, rather to domain-specific abilities. Movement is closely tied to music. Moving one or some parts of the body is a natural reaction to a rhythm. Much earlier on, Fraisse, Pichot, and Clairouin (1949) found that humans rock to a musical rhythm as young as one year of age, and we are able to tap along metronome beats by the age of three or four. This sensorimotor behavior is usually based on anticipation of when the next beat is coming, especially at slow tempi (Nozaradan, Peretz, & Keller, 2016; Steen, Jacoby, Fairhurst, & Keller, 2015), and the anticipation grows in accuracy with years of musical training (Nozaradan et al., 2016). Prediction of oncoming beats is also present in rhythms with tempo changes (Steen et al., 2015). The ability to synchronize to beats or changes in rhythm is conditioned by (1) motor limits, (2) the fractured memory of a slow sequence, (3) sensory difficulties at perceiving successiveness at fast tempi (Bartlett & Bartlett, 1959), and (4) the complexity of the rhythm (Fraisse & Repp, 2012). However, the adaptation is fast, since it can take as few as three consecutive taps accompanying each a consecutive beat in a novel rhythm to achieve a relative simultaneity (Fraisse & Repp, 2012). The rhythm structure affects the synchronization pattern, as some tempo speeds are easier to adjust to than others. When the interval between beats is on or under 1500 ms, the synchronization behavior tends towards anticipation of the next beat, while this anticipation rate starts to decrease with beat onset intervals of 1800 ms (Miyake, Onishi, & Poppel, 2004). From the previous investigations on rhythm, we can see that synchronization is basically prediction of the rhythm timings. The findings on the relationship between rhythmic abilities and linguistic competence suggest that this relationship might only surface in certain cases, and that transfer from modalities may affect how strong the relationship is. Some of the factors conditioning the existence of a connection between the two domains are atypical development, early stages of cognitive development, nature of L1, or linguistic area in L2 learning. Investigations from other auditory areas, such as speech anticipation, may clarify if and how rhythmic abilities and linguistic abilities are related. Other areas that might be helpful in establishing the nature of the rhythm-language relationship is taking into account how rhythm imbibes from visual and spatial cues. One of the main visual-spatial cues related to rhythm is movement. Movement, like language, is entrenched in our everyday life. We use movement to respond to virtually everything. When driving, we anticipate what other drivers or pedestrians are going to do and adjust accordingly by making foot, manual and head movements. When catching an object in a free fall, we anticipate (more or less successfully) the trajectory and the speed to move our arm and hand fast enough and try to grab the object before it hits the floor.

# References

Anvari, S. H., Trainor, L. J., Woodside, J., & Levy, B. A. (2002). Relations among musical skills, phonological processing, and early reading ability in preschool children. *Journal of Experimental Child Psychology*, *83*(2), 111–130.

Bartlett, N. R., & Bartlett, S. C. (1959). Synchronization of a motor response with an anticipated sensory event. *Psychological Review*, *66*(4), 203.

Beckman, M. E. (1986). *Stress and non-stress accent* (Vol. 7). Walter de Gruyter.

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

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).

Carr, K. W., White-Schwoch, T., Tierney, A. T., Strait, D. L., & Kraus, N. (2014). Beat synchronization predicts neural speech encoding and reading readiness in preschoolers. *Proceedings of the National Academy of Sciences*, *111*(40), 14559–14564.

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.

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.

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.

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

Degé, F., & Schwarzer, G. (2011). The effect of a music program on phonological awareness in preschoolers. *Frontiers in Psychology*, *2*, 124.

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.

Dilley, L. C. (2005). *The phonetics and phonology of tonal systems* (PhD thesis). Massachusetts Institute of Technology.

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.

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

Fraisse, P., Pichot, P., & Clairouin, G. (1949). Les aptitudes rythmiques. Etude comparée des oligophrenes et des enfants normaux. *Journal de Psychologie Normale et Pathologique*.

Fraisse, P., & Repp, B. H. (2012). Anticipation of rhythmic stimuli: Speed of establishment and precision of synchronization (1966). *Psychomusicology: Music, Mind, and Brain*, *22*(1), 84.

Fry, D. B. (1955). Duration and intensity as physical correlates of linguistic stress. *The Journal of the Acoustical Society of America*, *27*(4), 765–768.

Fry, D. B. (1958). Experiments in the perception of stress. *Language and Speech*, *1*(2), 126–152.

Fry, D. B. (1965). The dependence of stress judgments on vowel formant structure. In *Phonetic sciences* (pp. 306–311). Karger Publishers.

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.

Gordon, M., & Roettger, T. (2017). Acoustic correlates of word stress: A cross-linguistic survey. *Linguistics Vanguard*, *3*(1).

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

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.

Iversen, J. R., Patel, A. D., Nicodemus, B., & Emmorey, K. (2015). Synchronization to auditory and visual rhythms in hearing and deaf individuals. *Cognition*, *134*, 232–244.

Jentschke, S., Friederici, A. D., & Koelsch, S. (2014). Neural correlates of music-syntactic processing in two-year old children. *Developmental Cognitive Neuroscience*, *9*, 200–208.

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

Koelsch, S., & Siebel, W. A. (2005). Towards a neural basis of music perception. *Trends in Cognitive Sciences*, *9*(12), 578–584.

Lagrois, M.-É., Palmer, C., & Peretz, I. (2019). Poor synchronization to musical beat generalizes to speech. *Brain Sciences*, *9*(7), 157.

Lebedeva, G. C., & Kuhl, P. K. (2010). Sing that tune: Infants’ perception of melody and lyrics and the facilitation of phonetic recognition in songs. *Infant Behavior and Development*, *33*(4), 419–430.

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.

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

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.

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.

Martı'nez-Castilla, P., & Sotillo, M. (2014). Pitch processing in children with williams syndrome: Relationships between music and prosody skills. *Brain Sciences*, *4*(2), 376–395.

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.

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.

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.

Nozaradan, S., Peretz, I., & Keller, P. E. (2016). Individual differences in rhythmic cortical entrainment correlate with predictive behavior in sensorimotor synchronization. *Scientific Reports*, *6*, 20612.

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.

Patel, A. D., Wong, M., Foxton, J., Lochy, A., & Peretz, I. (2008). Speech intonation perception deficits in musical tone deafness (congenital amusia). *Music Perception: An Interdisciplinary Journal*, *25*(4), 357–368.

Pearce, J. (2005). Selected observations on amusia. *European Neurology*, *54*(3), 145–148.

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.

Salimpoor, V. N., Zald, D. H., Zatorre, R. J., Dagher, A., & McIntosh, A. R. (2015). Predictions and the brain: How musical sounds become rewarding. *Trends in Cognitive Sciences*, *19*(2), 86–91.

Sammler, D., Novembre, G., Koelsch, S., & Keller, P. E. (2013). Syntax in a pianist’s hand: ERP signatures of “embodied” syntax processing in music. *Cortex*, *49*(5), 1325–1339.

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?

Sereno, J. A., & Lee, H. (2015). The contribution of segmental and tonal information in mandarin spoken word processing. *Language and Speech*, *58*(2), 131–151.

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

Sluijter, A. M., & Van Heuven, V. J. (1996). Spectral balance as an acoustic correlate of linguistic stress. *The Journal of the Acoustical Society of America*, *100*(4), 2471–2485.

Sluijter, A. M., Van Heuven, V. J., & Pacilly, J. J. (1997). Spectral balance as a cue in the perception of linguistic stress. *The Journal of the Acoustical Society of America*, *101*(1), 503–513.

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.

Steen, M. M. van der, Jacoby, N., Fairhurst, M. T., & Keller, P. E. (2015). Sensorimotor synchronization with tempo-changing auditory sequences: Modeling temporal adaptation and anticipation. *Brain Research*, *1626*, 66–87.

Sun, Y., Lu, X., Ho, H. T., Johnson, B. W., Sammler, D., & Thompson, W. F. (2018). Syntactic processing in music and language: Parallel abnormalities observed in congenital amusia. *NeuroImage: Clinical*, *19*, 640–651.

Swaminathan, S., Schellenberg, E. G., & Venkatesan, K. (2018). Explaining the association between music training and reading in adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(6), 992.

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.

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.

Vickie, Y. Y., & Andruski, J. E. (2010). A cross-language study of perception of lexical stress in english. *Journal of Psycholinguistic Research*, *39*(4), 323–344.

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

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.

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.