Journal of Experimental Psychology:
Learning, Memory, and Cognition
2016, Vol. 42, No. 7, 1127–1139

0278-7393/16/$12.00 http://dx.doi.org/10.1037/xlm0000226

Listening Natively Across Perceptual Domains?

Alan Langus
SISSA—International School for Advanced Studies

Shima Seyed-Allaei
Institute for Research in Fundamental Sciences and University of Milan Bicocca

Ertuğrul Uysal
SISSA—International School for Advanced Studies and Universidade de Lisboa

Sahar Pirmoradian
SISSA—International School for Advanced Studies and University of Edinburgh

Caterina Marino
SISSA—International School for Advanced Studies

Sina Asaadi
Shahid Beheshti University of Medical Sciences

Ömer Eren
Boğaziçi University

Juan M. Toro
ICREA—Universitat Pompeu Fabra

Marcela Peña
Pontificia Universidad Catolica de Chile

Ricardo A. H. Bion
Stanford University

Marina Nespor
SISSA—International School for Advanced Studies

Our native tongue influences the way we perceive other languages. But does it also determine the way we perceive nonlinguistic sounds? The authors investigated how speakers of Italian, Turkish, and Persian group sequences of syllables, tones, or visual shapes alternating in either frequency or duration. We found strong native listening effects with linguistic stimuli. Speakers of Italian grouped the linguistic stimuli differently from speakers of Turkish and Persian. However, speakers of all languages showed the same perceptual biases when grouping the nonlinguistic auditory and the visual stimuli. The shared perceptual biases appear to be determined by universal grouping principles, and the linguistic differences caused by prosodic differences between the languages. Although previous findings suggest that acquired linguistic knowledge can either enhance or diminish the perception of both linguistic and nonlinguistic auditory stimuli, we found no transfer of native listening effects across auditory domains or perceptual modalities.

Keywords: native listening, cross-linguistic differences, transfer, perceptual grouping

Our native language influences the way we perceive foreign tongues. For example, upon President Clinton's first state visit to Japan, the Japanese people faced a dilemma. The Japanese language is particular because it has very few possible consonant clusters and it does not distinguish the phonemes *r* and *l*. Japanese speakers thus perceive *l*'s as *r*'s and hear illusory vowels between

This article was published Online First January 28, 2016.

Alan Langus, Neuroscience Area, SISSA—International School for Advanced Studies; Shima Seyed-Allaei, School of Cognitive Sciences, Institute for Research in Fundamental Sciences and Department of Psychology, University of Milan Bicocca; Ertuğrul Uysal, Neuroscience Area, SISSA—International School for Advanced Studies and Faculdade de Letras, Universidade de Lisboa; Sahar Pirmoradian, Neuroscience Area, SISSA—International School for Advanced Studies and Institute for Adaptive and Neural Computation, School of Informatics, University of Edinburgh; Caterina Marino, Neuroscience Area, SISSA—International School for Advanced Studies; Sina Asaadi, Functional Neurosurgery Research Center, Shahid Beheshti University of Medical Sciences; Ömer Eren, Department of Western Languages and Literatures, Boğaziçi University; Juan M. Toro, ICREA—Universitat Pompeu Fabra; Marcela Peña, Escuela de Psicologıa, Pontificia Universidad Catolica de Chile; Ricardo A. H. Bion, Department of Psychology, Stanford University; Marina Nespor, Neuroscience Area, SISSA—International School for Advanced Studies.

The research leading to these results has received funding from the European Research Council (ERC) under the European Union's Seventh Framework Programme (FP7/2007–2013)/ERC grant agreement nu 269502 (PASCAL) and the SISSA Young Researcher Grant.

Correspondence concerning this article should be addressed to Alan Langus, Language, Cognition and Development Laboratory, Cognitive Neuroscience Sector, SISSA—International School for Advanced Studies, via Bonomea 265, Trieste 34136, Italy. E-mail: langus@sissa.it

adjacent consonants (Goto, 1971; Dupoux et al., 1999). However, *Biru Kurinton*, the adaptation of the President's name, sounded so much like the Japanese word *kurikinton* (translated as *sweet potato*) that many wondered who this *Biru sweet potato* was. These perceptual effects are collectively called *native listening* (see Cutler, 2012 for an overview). Native listening effects emerge whenever we encounter aspects of foreign languages that are absent in our own. Depending on our native language and the language we are listening to, they can be found at every level of linguistic representations from phonemes (Werker et al., 1981; Dupoux et al., 1999; Dupoux et al., 2008), through to syllables (Cutler et al., 1983; Cutler et al., 1986) and words (Toro et al., 2011). They can even occur with suprasegmental cues like lexical stress (Dupoux et al., 1997; Dupoux, Peperkamp, & Sebastian, 2001; Peperkamp & Dupoux, 2002) and with the interpretation of intonational contours (Gussenhoven & Chen, 2000; Makarova, 2001). Because native listening so systematically distorts our perception of foreign languages, is it possible that the knowledge of our mother tongue also influences our perception of nonlinguistic sounds?

The investigation of native language effects on nonlinguistic sounds has primarily focused on music. As the cortical regions responsible for processing spoken language and musical melodies engage some of the same neural regions (Patel, 2003), damage in these regions can cause highly correlated disadvantages in the processing of both speech and music. For example, amusic individuals may experience difficulties in perceiving the prosody of spoken language (Patel, Peretz, Tramo, & Labreque, 1998), and aphasic individuals can have difficulties in perceiving the structure of music (Patel, 2008). However, despite converging evidence for the shared neural basis for the processing of music and spoken language (Patel, 2003), native language effects in music perception in healthy adults remain controversial. For example, Mandarin speakers misidentify certain tone combinations in nonspeech pitch contours that they assimilate to the lexical tones used in their native language (Bent, Bradlow & Wright, 2006). In contrast, speakers of Cantonese, also a tonal language, appear to be better at processing musical pitch than speakers of French and English, (Wong, Ciocca, Chan, Ha, Tan, & Peretz, 2012), but do not differ from English and French speakers in determining whether musical melodies are mistuned or off-beat (e.g., violating rhythm). This suggests that linguistic experience with tones could increase listeners' attention to low-level acoustic cues and facilitate discrimination of such acoustic properties as pitch.

There is some evidence suggesting that the language we speak may also influence higher-level musical perception. One of the few known cases where speakers of different tongues have categorically different musical experiences when listening to exactly the same tone combinations is the tritone paradox (Deutsch, 1986). The tritone paradox arises when two tones related by a half-octave (a tritone) are played sequentially. Some subjects listening to such tones report hearing an ascending pattern, but others a descending pattern, depending on their native language or dialect. For example, subjects born in California and south England systematically report the opposite percepts (Deutsch, 1991, 2013). Whether a listener perceives an ascending or a descending pattern is correlated with the pitch range of the listeners' speech (Deutsch, North, & Ray, 1990). Because speakers adjust their pitch range to the pattern of pitch ranges of the language spoken in their community, the tritone paradox can cause speakers of different languages and even dialects of the same language to hear exactly opposite patterns when listening to tritone intervals (Deutsch, 2013). To what extent is the tritone interval and the variation in pitch perception by speakers of different languages thus a symptom of how our native language affects our perception of nonlinguistic sounds?

The Iambic–Trochaic Law

The knowledge of our native language can distort our perception of nonlinguistic auditory stimuli if speech and nonspeech auditory perception share common mechanisms and/or representations that are domain general. One such domain general mechanism is the Iambic–Trochaic Law (ITL). An *iamb* is a group of two elements where the strong element follows the weak one and a *trochee* a group where the strong element precedes the weak one. The ITL thus describes the perceptual phenomenon according to which humans are biased to group elements alternating in prominence either iambically or trochaically depending on the physical properties that signal prominence. The ITL was first formally proposed in the domain of music (Bolton, 1894). Consequent research in the perception of rhythm suggests that iambic/trochaic grouping of pure tones can be caused by a number of different stimulus properties. For example, prominence alternating in intensity tends to lead listeners to group pure tones in trochaic (strong-weak) pairs (Fraisse, 1956; Jones, 1981), as does prominence alternating in pitch (Jones, 1981). Conversely, prominence alternating in duration tends to lead listeners to group pure tones in iambic (weak-strong) pairs (Woodrow, 1951). This leads to the following formulation of the ITL:

- Elements alternating in pitch and/or intensity are grouped into trochees with the strong element (with higher pitch/intensity) preceding the weak one.
- Elements alternating in duration are grouped into iambs with the strong element (with the longer duration) following the weak one.

The ITL has been shown to be applicable to two separate levels of rhythm in natural language phonology. The ITL accounts for the allocation of word secondary stress, that is, if language, in forming feet, systematically puts heavy syllables at the right edge, then the rightmost syllable in the foot will be marked mainly by longer duration, whereas if a language ignores syllable weight in forming feet, then it will form trochaic feet by stressing the leftmost syllable through intensity (Hayes, 1995).

Importantly, the ITL describes the alternation of prominence but lexical stress does not alternate. For example, both the word *cat* and the word *supercalifragilisticexpialidocious* only have one lexical stress. Because the assignment of secondary stress(es) is dependent on the number of feet within a word, the boundaries of perceptual groups of iambs or trochees do not necessarily correspond to word boundaries. This has caused some confusion in the literature because in some languages, like English, where lexical stress is often word initial (sometimes called trochaic) and many words are just one metrical foot long, listeners may acquire a segmentation strategy that assigns word boundaries according to lexical stress (Cutler & Norris, 1988; Cutler & Carter, 1987; Cutler, 1990) and yields trochaic feet (Jusczyk, Cutler, Redanz, 1993). However, because not all feet in English carry primary lexical stress and stress does not always fall on the initial syllable (e.g., *delicious*), even in English the segmentation strategy based on lexical stress does not adhere to the ITL (cf. Gerken, 1994). That segmenting and grouping syllables into iambs/trochees are two different processes can be seen in the behavior of French speakers. If familiarized with streams of syllables segregated by pauses, which alternate either in intensity or duration, French speakers group these syllables into trochees or iambs according to the ITL (Hay & Diehl, 2007). However, French speakers familiarized with continuous syllable streams where the syllables alternate either in pitch, intensity or duration, segment the stream into stress initial chunks with pitch/intensity but fail to segment the stream into stress final chunks with duration (Bhatara et al., 2013), a failure that occurs also with sequences of highly variable sequences of connected instrumental sounds that mimic speech (Bhatara et al., 2015). Because the ITL proposed by Hayes (1995) does thus not define prominence at any unit other than the foot, it is unlikely to significantly contribute to language perception and acquisition beyond the level of metrical feet.

It is not at the level of feet that the ITL should be at its best in aiding the listener in speech perception. In spoken languages, the ITL has also been shown to characterize prominence at the Phonological Phrase (PP) level, where pitch in addition to intensity characterizes trochaic prominence. The PP extends from the left edge of a syntactic phrase to the right edge of its head in head-complement languages; and from the left edge of a head to the left edge of its syntactic phrase in complement-head languages (Nespor & Vogel, 2007). Languages with the basic object-verb (OV) word order where the head of the phrase follows its complements, such as Turkish and Japanese, mark prominence mainly trochaically through pitch and intensity on the stressed syllable of the first word of the phonological phrase. Languages with the basic verb-object (VO) word order where the head of the phrase precedes its complements, like English and Italian, mark prominence mainly iambically through duration on the stressed syllable of the last word of phonological phrases (Nespor, Shukla, van de Vijver, Avesani, Schraudolf, & Donati, 2008). Importantly, because phonological phrases never straddle word boundaries (Nespor & Vogel, 2007), the ITL also helps listeners to group segmented words into Phonological Phrases (Langus et al., 2012), and may even cue the basic word order of the language of exposure (Nespor et al., 2008).

A growing body of experimental evidence from speakers of VO languages suggests that the ITL guides the perception of rhythm across several perceptual domains. English- and French-speaking adults group pure tones and syllables iambically if these alternate in duration and trochaically if these alternate in intensity (Hay & Diehl, 2007). Italian-speaking adults group syllables alternating in duration iambically and syllables alternating in pitch trochaically (Bion et al., 2011). Italian-speaking participants group even visual events alternating in temporal frequency/intensity trochaically and visual events alternating in duration iambically (Peña, Bion, & Nespor, 2011). Studies with speakers of VO languages thus suggest that the ITL is a domain general perceptual mechanism. However, Japanese-speaking adults, the only speakers of an OV language tested so far on the ITL, show no consistent preference for pure tones alternating in duration (Kusumoto & Moreton, 1997; Iversen, Patel, & Ohgushi, 2008). Because Japanese is an OV language and marks prominence trochaically in phonological phrases mainly through pitch and intensity (Nespor et al., 2008), Japanese-speakers inability to group tones alternating in duration may emerge as a result of their native language. The ITL is therefore interesting for studying native listening across perceptual domains because the word order of the listener's native language may influence their biases for grouping elements alternating in prominence into either iambs or trochees.

There is no evidence that speakers of any given language can group different types of stimuli—syllables, nonlinguistic auditory stimuli as well as visual events—into trochees if these alternate in pitch/intensity and into iambs if these alternate in duration. Furthermore, participants from linguistic backgrounds who are known to violate the ITL over nonlinguistic tones (e.g., Japanese speakers) have never been tested with linguistic and visual stimuli. Cross-linguistic comparisons of participants' performance with the ITL in different perceptual domains are therefore necessary not only for verifying whether universal grouping biases for iambs and trochees exist in the first place. Understanding whether these violations emerge systematically in speakers of languages that share specific prosodic characteristics can also indicate whether the acquired knowledge of our native language can override basic perceptual biases such as the ITL and maybe even influence the way perceive rhythm across different perceptual domains. We thus compare how speakers of OV and VO languages group elements alternating in pitch and duration in different perceptual domains that include linguistic stimuli, nonlinguistic auditory, and visual stimuli.

Experiment 1: Perceptual Grouping of Syllable Sequences

In Experiment 1, we tested whether native language effects emerge in grouping linguistic sounds (syllable pairs) that alternate in either pitch or duration. We use pitch instead of intensity because the perception of pitch in music, but not that of intensity, appears to be influenced by the native language of the listeners (Bent, Bradlow & Wright, 2006; Wong, Ciocca, Chan, Ha, Tan, & Peretz, 2012). Native listening effects in the perception of rhythm signaled through pitch would thus be likely to transfer to other perceptual domains. Pitch is also useful because a failure in grouping syllables alternating in pitch trochaically can directly be related to a specific level in the prosodic hierarchy, because pitch only signals trochaic grouping in phonological phrases (Nespor et al., 2008). All studies testing grouping preferences for syllables alternating in pitch/intensity or duration in adults have been carried out on speakers of VO languages such as English, Italian, and French (Hay & Diehl, 2007; Bion et al., 2011), that signal prominence in phonological phrases iambically through duration (Nespor et al., 2008). However, it is unknown how speakers of OV languages, which signal phrasal prominence primarily through pitch and intensity, would group syllables alternating in duration. We therefore tested Italian- (VO), Persian- (OV), and Turkish-speaking (OV) adults. In Italian phrasal stress is iambic (signaled mainly through duration at the end of the phonological phrase) and in Persian as well as in Turkish it is trochaic (signaled mainly though pitch and intensity at the beginning of the phonological phrase) (Nespor et al., 2008; Nespor & Vogel, 2007; Windfuhr, 1979). In Italian, word primary stress is most often in word penultimate position and it is signaled mainly through duration (Nespor, 1993; Ordin & Nespor, 2013); in Persian word primary stress, signaled mainly through intensity, falls on the rightmost stressable syllable

(Ayat Hosseini, 2014); and in Turkish word primary, stress is word final and is also signaled mainly through intensity (Kornfilt, 1997). Finally, word secondary stress is trochaic in Italian (Nespor, 1993), it is absent in Persian (Kahnemuyipour, 2003), and it is in word final position in morphologically complex words in Turkish (Revithiadou et al., 2006; though see Kabak & Vogel, 2001 for arguments against secondary stress in Turkish). The most common realizations of prominence in the three languages are summarized in Table 1.

Based on previous findings, we predict that Italian speakers should perceive elements alternating in prominence according to the ITL (Bion et al., 2011). In Italian, pitch does not signal prominence at the level of feet, words or phonological phrases. Because Italian-speaking adults and infants have been found to group syllables alternating in pitch into trochees (high-low), the trochaic bias with pitch may emerge even without exposure to linguistic prominence signaled through pitch. Although Turkish and Persian use pitch to signal prominence both in phonological phrases and in words, in both languages phonological phrase prominence is phrase initial and lexical stress is word final. If Turkish and Persian speakers interpret the prominence signaled through pitch as primary lexical stress, then listeners should prefer iambic (low–high) syllable pairs to trochaic (high–low) ones. Turkish and Persian speakers' performance can therefore determine whether grouping biases for syllables alternating in pitch emerge as the result of lexical stress or phonological phrase prominence. Conversely, while duration signals both phonological phrase level prominence and lexical stress in Italian, it does not signal prominence in Persian and Turkish. If the iambic bias for elements alternating in duration is universal, then Italian, Turkish and Persian speakers should follow the ITL and group syllables alternating in duration as short-long. Previous findings with Italian-learning infants suggest that the ITL grouping for duration may emerge as a result of linguistic experience (Bion et al., 2011). Speakers of languages like Turkish and Persian, who have no experience with prominence signaled through duration, could therefore also differ from Italians and show no iambic bias for syllables alternating in duration.

Method

Participants. We recruited 24 Italian-speaking adults from Trieste (Italy), 24 Persian-speaking adults from Tehran (Iran), and 24 Turkish-speaking adults from Istanbul (Turkey; 42 females, mean age 22 ± 4 years). All participants were monolingual and reported no auditory, visual, or language-related problems.

Stimuli. The stimuli of Experiment 1 consisted of two familiarization streams. Both streams contained the same 10 adjacent syllables (/pa su tu ke ma vi bu go ne du/) separated by 100-ms long pauses and repeated in the same order 20 times. The phonemes were 180-ms long and had a fundamental frequency of 180 Hz. In one of the familiarization streams prominence was signaled with fundamental frequency. The stress was instantiated by increasing the fundamental frequency of the central portion of the vowel of the odd syllables to 400 Hz. In the second familiarization stream prominence was signaled with duration. The stress was instantiated by increasing the duration of the vowel of every odd syllable to 400 ms. The familiarization streams were faded-in and -out by increasing and decreasing amplitude ramps in the first and last 5-s periods. The two adjacent syllables forming a group in the pitch stream (i.e., pa-su, tu-ke, ma-vi, bu-go, ne-du) were straddling group boundaries in the duration stream (i.e., su-tu, ke-ma, vi-bu, go-ne, du-pa): The five conforming syllable pairs that were the correct answers in one condition were thus the nonconforming syllable pairs (the incorrect answers) in the other condition. All the test stimuli were prosodically flat and had a constant fundamental frequency of 180Hz and phoneme-duration of 180ms. We know of no words in any of the languages that would match these syllable pairs. The stimuli were synthesized with the IT4 MBROLA database using PRAAT (Boersma, 2001) and MBROLA (Dutoit et al., 1996).

Procedure

Participants were tested in a silent room with stimuli presented through Sony MDR-XB700 overear headphones. Participants were randomly assigned either to the pitch condition or to the duration condition so that half of the participants from each language group participated in the pitch condition and the other half in the duration condition. Participants were instructed to listen carefully because they would be queried about the syllable pairs they heard during the familiarization phase. After listening to the familiarization stream, participants were auditorily presented with the test stimuli consisting of one syllable pair at the time. Participants had to indicate yes/no to whether they had heard the syllable pairs in the given order during familiarization. Because all 10 syllable pairs

Table 1
Dominant Phonological Phrase Prominence, Word Primary Stress and Word Secondary Stress Patterns in Italian, Persian, and Turkish

Language	Phonological phrase (PP)	Word primary stress	Word secondary stress
Italian	Iambic		Trochaic
	PP final and signaled primarily through duration	On one of the last 3 syllables. Mostly penultimate. Signaled primarily through duration.	Foot initial and signaled primarily through intensity
Persian	Trochaic		
	PP initial and signaled primarily through pitch and intensity	Right-most stressable syllable. Signaled through pitch and intensity.	No secondary stress on words
Turkish	Trochaic	Word final	Word final
	PP initial and signaled primarily through pitch and intensity	Signaled through pitch and intensity.	Occurs only in compounds, on the word that does not bear the compound's primary stress. It can occur both on the first or the second word in the compound. Signaled through intensity.

presented during test were adjacent during the familiarization—but half of the test syllable pairs (half of the trials) followed iambic and the other half trochaic grouping—any consistent choice would be extremely strong evidence for perceptual grouping of linguistic sounds. Each test item was presented twice and the order of the test items was randomized across participants. The experiment was controlled by PsyScope X software (http://psy.ck.sissa.it/).

Results and Discussion

The results of Experiment 1 are shown in Figure 1. Because calculating the overall percentage of correct answers does not take into account possible biases participants may have in answering (e.g., a strong preference toward one response), we compared participants' sensitivity and bias using the signal detection theory (SDT; Tanner & Swets, 1954). Participants' responses were coded as hits (saying *yes* to syllable pairs that adhered to ITL), misses (saying *no* to pairs that adhered ITL), correct rejections (saying *no* to pairs that violated ITL) and false alarms (saying *yes* to pairs that violated ITL). We used d' (also called sensitivity) that measures participants' sensitivity in discriminating syllable pairs that adhered to the ITL during the familiarization from syllable pairs that did not. d' values above 0 indicate discriminability and values around 0 chance performance.

Because the data is binomially distributed, that is, participants could only respond yes/no, we analyzed the data using general linearized mixed models (GLMM). The GLMMs are better suited for investigating binomial data than classical ANOVAs because they can take into account the binomial distribution of the data and additionally also allow for modeling possible random effects arising from individual differences between subjects and experimental items. In the GLMM framework the traditional SDT model is equivalent to a probit regression (DeCarlo, 1998, 2010) where participants' responses (yes/no) are predicted from the true value of the test item (adheres to ITL/does not adhere to ITL) through a probit link function. The intercept of the probit regression reflects any overall bias in answering yes/no (criterion $c = -1^{*}$intercept), and the slope measures participants' sensitivity to syllable pairs adhering to the ITL (d' = slope). Interactions with other variables such as participants' native language and the experimental condition can be included as fixed effects. Significant interactions between true value of the test item and these fixed parameters would indicate that native language and condition affect participants' ability to discriminate syllable pairs that adhere to the ITL from those that violate it. We included participant and experimental items as random effects.

All models were implemented in the R statistical environment (R Development Core Team, 2012) using the *glmer* function from the *lme4* package for binomial data (Bates, 2005). Because we were interested in the d' scores, we used the 'probit' link-function and the Laplacian approximation (cf. DeCarlo, 1998, 2010). The data was coded in a $n \times 6$ matrix that contained participants' responses (coded as *yes* = 1/*no* = 0), whether the test item adhered to the ITL (*adhered/did not adhere to ITL* = .5/-.5), subject number (Subno), item number (itmnr), participants' native language (Italian/Persian/Turkish), and their condition (pitch/duration). Significance of the main effects and their interactions was determined by comparing simpler models to more complex models by including parameters (e.g., native language or experimental condition) one by one. Model selection revealed a significant main effect of discrimination between syllable pairs adhering/not adhering to ITL, $\chi^2(df = 6) = 288.32$, $p < .001$; a significant interaction between the test items' adherence to ITL and participants' native language, $\chi^2(df = 6) = 108.49$, $p < .001$; a significant interaction between test items' adherence to ITL and experimental condition, $\chi^2(df = 5) = 218.26$, $p < .001$; as well as a significant interaction between test items' adherence to ITL, participants' native language and experimental condition, $\chi^2(df = 7) = 289.26$, $p < .001$. To investigate the three-way interaction, we analyzed the two experimental conditions (syllables alternating in pitch and in duration) separately.

Participants grouped syllables alternating in pitch into trochees regardless of their native language. We found a significant main effect of discrimination between test syllable pairs adhering/not adhering to the ITL, $\chi^2(df = 1) = 12.33$, $p < .001$, but no interaction between test items' adherence to ITL and participants' native language, $\chi^2(df = 2) = 3.78$, $p = .151$: Italian-speaking participants accepted trochaic (high–low) and rejected iambic (low–high) syllable pairs above chance ($d' = 2.47$, $SE = .51$, $z = 4.87$, $p < .0001$), just like Persian- ($d' = 1.94$, $SE = .49$, $z = 3.97$, $p < .0001$) and Turkish-speaking adults ($d' = 1.65$, $SE = .48$, $z = 3.42$, $p < .001$). However, grouping of syllables alternating in duration depended on participants' native language. Even though we found no significant main effect of discrimination between test syllable pairs adhering/not adhering to the ITL, $\chi^2(df = 1) = 1.38$, $p = .240$, we did find a significant interaction between test items' adherence to ITL and participants' native language, $\chi^2(df = 2) =$

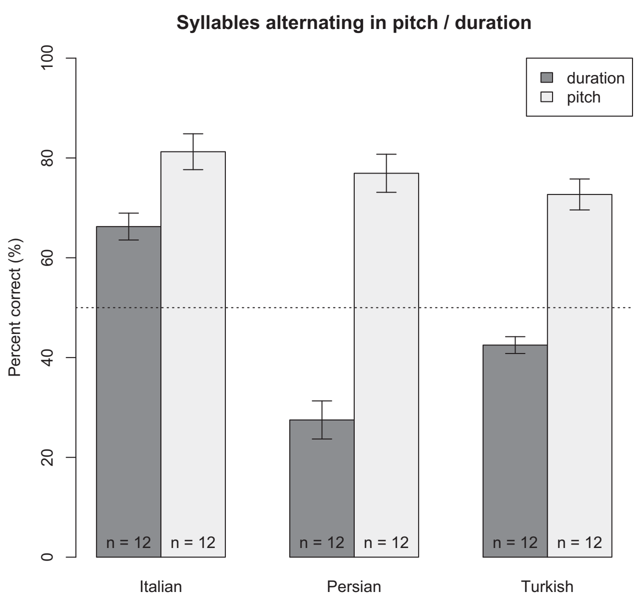

Figure 1. Italian-, Turkish-, and Persian-speaking participants' grouping preferences of syllable sequences alternating in fundamental frequency and in duration. Y-axis denotes the percent of Iambic-Trochaic grouping: (a) correctly accepting high–low (strong–weak) syllable pairs and rejecting low–high (weak–strong) syllable pairs in the fundamental frequency condition; and (b) correctly accepting short–long (weak–strong) and rejecting long–short (strong–weak) syllable pairs in the duration condition. Error bars denote standard error. The asterisk (*) denotes $p < .05$ against chance.

91.98, $p < .0001$. The results revealed an iambic bias in Italian-speaking participants, but a trochaic bias in Persian- and Turkish-speaking adults. Although Italian-speaking participants accepted iambic (short–long) and rejected trochaic (long–short) syllable pairs significantly above chance ($d' = 1.10$, $SE = .01$, $z = 308.9$, $p < .0001$), Persian- ($d' = -1.38$, $SE = .30$, $z = -4.70$, $p < .0001$) and Turkish-speaking participants ($d' = -.42$, $SE = .10$, $z = -116.5$, $p < .001$) accepted trochaic (long-short) and rejected iambic (short-long) syllable pairs above chance.

The results of Experiment 1 replicate previous findings with speakers of VO languages by showing that Italian participants group syllables alternating in pitch trochaically (strong–weak/high–low) and syllables alternating in duration iambically (weak–strong/short–long). Also speakers of OV languages—in our case Turkish and Persian—show a robust trochaic bias when grouping syllables alternating in pitch. The trochaic bias does not only work over syllables alternating in intensity (Hay & Diehl, 2007), but it emerges cross-linguistically also over syllables that alternate in pitch. Pitch is the main cue of trochaic prominence in spoken language at the phonological phrase level only in OV languages (Nespor et al., 2008) and it never signals trochaic prominence at the level of feet (Hayes, 1995). Cross-linguistic evidence for a trochaic bias with pitch thus supports the idea that grouping syllables into trochees with pitch is a universal in language perception. This conclusion is further supported by the fact that in VO languages, such as Italian, pitch does not systematically signal trochaic prominence at any level of the prosodic hierarchy. Italian-speaking participants could thus not acquire the bias to group syllables alternating in pitch into trochees from their native language.

Our results also show that robust native language effects emerge in grouping syllables alternating in duration. Turkish- and Persian-speaking adults did not only fail to group syllables alternating in duration iambically (short–long), but they grouped them trochaically instead (long–short). Neither language uses duration to systematically signal prominence at the phrasal level. Prominence in phonological phrases in Persian and Turkish is trochaic (signaled through pitch and intensity on the word at the left edge of the phonological phrase), but word level stress is word final in both languages and word secondary stress is leveled in connected speech in Persian and rare in Turkish (Kabak & Vogel, 2001; Kahnemuyipour, 2003). The trochaic (strong–weak/long–short) bias in Persian- and Turkish-speaking participants in grouping syllables alternating in duration could thus not emerge from experience with duration as a cue to prominence in participants' native language. Instead, the results suggest that Turkish and Persian participants must have interpreted the prominence in the familiarization stream not in terms of the specific cue (e.g., short–long/long–short) but rather through the location of prominent syllable in relation to the nonprominent syllable (strong–weak rather than weak-strong). Because in Persian and Turkish the strong-weak rhythm can only be found at the level of phonological phrases, participants must have assimilated the prominence signaled through duration to the trochaic phrasal rhythm typical of their native languages. Rather than failing to group syllables alternating in duration because Turkish and Persian do not signal prominence with duration, Turkish- and Persian-speaking participants actually listened to syllables alternating in duration natively by relying on the representations of trochaic phonological phrase rhythm of their native language.

Why do Persian- and Turkish-speaking adults fail to show an iambic bias for syllables alternating in duration, whereas Italian-speaking adults group syllables alternating in pitch trochaically even though their native language does not signal trochaic prominence with pitch? Studies in language acquisition with young infants suggest that the reason why duration causes native listening effects and pitch does not may have a developmental origin. For example, young Italian-learning (VO) and Japanese-learning (OV) infants group syllables alternating in pitch trochaically regardless of the language they are learning, but fail to group syllables iambically even when they are learning a VO language such as Italian (Bion et al., 2011; Yoshida et al., 2010). This suggests that during cognitive development the trochaic bias emerges earlier than the iambic bias. This is also found with intensity in English-learning infants who show a trochaic bias earlier than the iambic bias (Hay & Saffran, 2012). Together with the results of Experiment 1, where the trochaic bias emerged even in participants who have no experience with trochaic rhythm signaled through pitch, this suggests that the trochaic bias may not require linguistic experience. The iambic bias to group syllables alternating in duration does not only appear later in development: It remains absent in Persian- and Turkish-speaking listeners whose native language does not have iambic rhythm. The data thus suggest that the perception of linguistic rhythm in adult listeners is governed both by universal and by language specific knowledge: syllables alternating in pitch are universally grouped into trochees, but the grouping of syllables alternating in duration depends on language specific knowledge.

Experiment 2: Perceptual Grouping of Sine-Wave Speech

Experiment 2 investigates whether participants can listen natively across perceptual domains. We tested whether the native listening effects we observed with OV-language speakers (e.g., Turkish and Persian) in Experiment 1 generalize to nonlinguistic stimuli. There is some evidence that native listening effects in the ITL may emerge also with nonlinguistic sounds. Although Japanese and English speakers group pure tones alternating in intensity trochaically, Japanese speakers do not show any bias to for tones alternating in duration (Kusumoto & Moreton, 1997; Iversen, Patel, & Ohgushi, 2008). Although it is unknown how speakers of Japanese, an OV language, would group syllables alternating in prominence, the results of Experiment 1 suggest that the failure to group tones iambically could emerge because the prosody of OV-languages does not signal prominence in phonological phrases through duration. As a consequence, also Turkish and Persian speakers should fail to show an iambic bias for nonlinguistic auditory stimuli alternating in duration.

As opposed to previous studies that have investigated the ITL with nonlinguistic stimuli (Kusumoto & Moreton, 1997; Iversen, Patel, & Ohgushi, 2008), we decided not to use pure tones. It is difficult to equate pure tones to syllables because syllables in Experiment 1 were composed of a consonant–vowel pair where only the vowel carried the prominence (high or low/short or long). Furthermore, when processing spoken language, prosodic cues (e.g., prominence) can be dissociated from the identity of segmental units (e.g., syllables). For example, when listening to a stressed syllable and listening to same syllable when it is unstressed, the listener is still capable of recognizing that (s)he is listening to exactly the same syllable. Because Experiment 1 relied on pitch, rather than intensity, to signal prominence over syllables, we can no longer use pure tones to study the trochaic bias over nonlinguistic stimuli. The fundamental frequency defines the identity of pure tones and it can therefore not be dissociated from the "prosodic" characteristics of pure tones in a way intensity would: that is, changing the fundamental frequency of a tone changes it into a different tone. While it would thus still possible to study whether tone sequences alternating in fundamental frequency would be grouped into high–low or low–high pairs, it would not be possible to determine whether a trochaic (high–low) bias would also facilitate the memory for certain tone pairs over others.

To determine whether the ITL can also facilitate the memory for nonlinguistic auditory stimuli, as it does for spoken language, it is important to test participants with acoustic stimuli where the alternation of pitch and duration do not directly define the identity of the segmental units. In Experiment 2, we converted the speech stimuli used in Experiment 1 into sine-wave analogues of speech. Sine-wave speech is synthesized by replacing three formants in speech with sinusoids. This transformation eliminates the phonetic information but preserves the overall complexity of spoken language, including the frequency and amplitude variations that define the prosody of spoken language (Remez, Rubin, Pisoni, & Carrell, 1981). The conversion of the stimuli of Experiment 1 into sine-wave speech results in acoustically different sine-wave speech chunks, the memory of which can be tested similarly to the syllable pairs in Experiment 1. Sine-wave analogues of speech may be particularly useful in testing the ITL because despite the acoustic similarities to spoken language, untrained and naïve listeners tend to perceive sine-wave speech nonlinguistically (Remez et al., 1981; Möttönen et al., 2006; see also Lee, & Noppeney, 2011). In fact, sine-wave speech is likely to be processed as linguistic input only if it is synthesized from linguistic utterances that have meaning (e.g., actual words) or if participants are informed beforehand that they are hearing linguistic stimuli (Remez et al., 1981; Remez et al., 2011). Sine-wave analogues of the stimuli used in Experiment 1, that were nonsense sequences of syllables, are likely to be processed nonlinguistically if participants remain unaware that the stimuli could be in linguistic in nature. This may enable us to determine whether alternation in pitch and duration facilitate memory for specific iambic/trochaic pairs also in more complex nonlinguistic stimuli.

Method

Participants. We recruited 24 Italian-speaking adults from Trieste (Italy), 24 Persian-speaking adults from Tehran (Iran), and 24 Turkish-speaking adults from Istanbul (Turkey) (42 females, mean age 22 ± 4 years). All participants were monolingual and reported no auditory, visual, or language related problems. None of the participants had participated in Experiment 1.

Stimuli and procedure. The stimuli of Experiment 2 were synthesized from the stimuli of Experiment 1. To make the stimuli of Experiment 3 as comparable as possible to the stimuli of Experiment 1, we used the Praat (Boersma & Weenink, 2009) script that fits sinusoids to changes in resonance within each syllable, creating sine-wave equivalents of speech around three formants (www.lifesci.sussex.ac.uk/home/Chris_Darwin/Praatscripts/SWS). This script uses a formant tracker to detect the formant frequencies found in speech stimuli and then replaces target formants with sinusoids by positioning time-varying sine-waves at the center frequencies of the three lower formants of the speech tokens. The transformation does not only preserve the exact temporal structure of the stimuli of Experiment 1, but also the frequency and amplitude variations of the original stimuli. For studies that rely on the dual nature of sine-wave speech—that is, that it can be perceived both as speech and as nonspeech—it is preferable to synthesize the stimuli manually. However, because we are not interested in participants perceiving the stimuli as speech, the automatic conversion may be sufficient to degrade the speech stimuli to the extent that they are no longer perceived linguistically, while retaining the complexity of spoken language. This resulted in a sinewave analogue of the familiarization stream of Experiment 1 and sinewave analogues of the prosodically flat test items of Experiment 1. The procedure of Experiment 2 was identical to that of Experiment 1, with the exception that participants at the end of the experiment were asked whether they had noticed that the stimuli were speech-like.

Results and Discussion

The results of Experiment 2 are shown in Figure 2. The procedure for analyzing the results was identical to that of Experiment

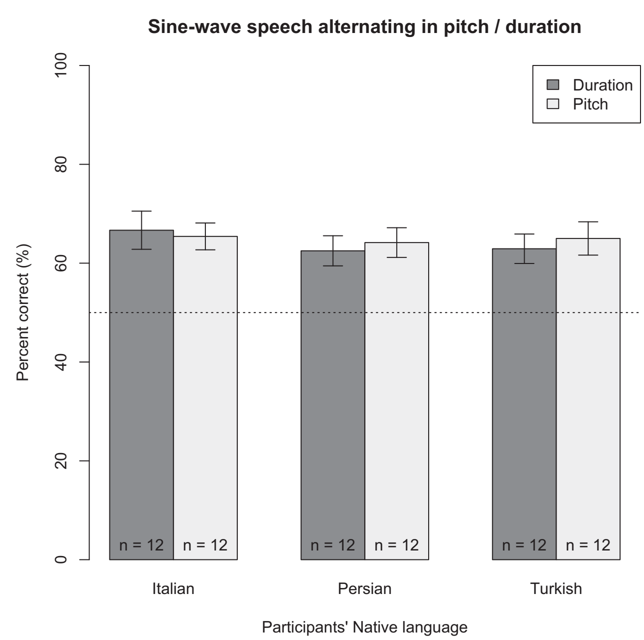

Figure 2. Italian-, Turkish-, and Persian-speaking participants' grouping-preferences of sine-wave sequences alternating in fundamental frequency and in duration. Y-axis denotes the percent of Iambic-Trochaic grouping: (a) correctly accepting high–low (strong–weak) sine-wave chunk pairs and rejecting low–high (weak–strong) sine-wave chunk pairs in the fundamental frequency condition; and (b) correctly accepting short–long (weak–strong) and rejecting long–short (strong–weak) sine-wave chunk pairs in the duration condition. Error bars denote standard error. The asterisk (*) denotes $p < .05$ against chance.

1. The results revealed that sine-wave speech chunks alternating in pitch or duration are grouped according to the ITL regardless of participants' native language. Model selection revealed a significant main effect of discrimination between test sine-wave speech chunk pairs adhering/not adhering to the ITL, $\chi^2(df = 6) = 12.61$, $p = .049$, but no interaction between test items' adherence to ITL and participants' native language, $\chi^2(df = 6) = 1.606$, $p = .952$ or experimental condition, $\chi^2(df = 5) = 2.509$, $p = .775$. Participants grouped sine-wave speech chunks alternating in pitch into trochees regardless of their native language. We found a significant main effect of discrimination between test items' adhering/not adhering to the ITL, $\chi^2(df = 1) = 8.848$, $p < .01$, but no interaction between test items' adherence to ITL and participants native language, $\chi^2(df = 2) = 0.189$, $p = .909$. Italian-speaking participants accepted trochaic (high–low) and rejecting iambic (low–high) sine-wave speech chunks above chance ($d' = 1.123$, $SE = .35$, $z = 3.17$, $p = .002$), just like Persian- ($d' = 1.055$, $SE = .36$, $z = 2.96$, $p = .003$) and Turkish-speaking adults ($d' = 1.196$, $SE = .36$, $z = 3.28$, $p = .001$). Participants grouped also sine-wave speech chunks alternating in duration into iambs regardless of their native language. We found a significant main effect of discrimination between test items' adhering/not adhering to the ITL, $\chi^2(df = 1) = 7.96$, $p < .005$, but no interaction between test items' adherence to ITL and participants' native language, $\chi^2(df = 2) = 1.302$, $p = .521$. Italian-speaking participants accepted iambic (short-long) and rejected trochaic (long-short) sine-wave speech chunks above chance ($d' = 1.12$, $SE = .34$, $z = 3.209$, $p = .001$), just like Persian- ($d' = 0.86$, $SE = .34$, $z = 2.47$, $p = .0135$) and Turkish-speaking adults ($d' = 0.87$, $SE = .35$, $z = 2.50$, $p = .0125$).

Italian, Persian, and Turkish speakers grouped sine-wave speech chunks iambically if prominence was signaled through duration and, trochaically if it was signaled through pitch. Even though sine-wave speech preserves the complexity and the prosodic structure of spoken language, the native listening effects we found in Turkish and Persian speakers in Experiment 1 did not emerge in Experiment 2. This suggests that the native listening effects observed in the perception of rhythm in spoken-language do not necessarily transfer to nonlinguistic stimuli. This is interesting because, naïve listeners often fail to realize that sine-wave speech is linguistic in nature (Remez et al., 1981, 2011), as did all our participants. Our results thus also show that even after a 3-min exposure to sine-wave analogues of speech synthesized from nonsense syllable strings, Italian-, Persian-, and Turkish-speaking adult participants do not perceive these strings linguistically. Importantly, participants recognized the pairs of sine-wave analogues of spoken syllables significantly above chance, showing that the iambic-trochaic prominence does facilitate memory for grouped sine-wave chunks. This is further supported by the fact that the test items did not have any prominence: that is, they were synthesized from prosodically flat syllable pairs. This suggests that the ITL does not only cause participants to group elements alternating in prominence into pairs but also facilitates the memory for those pairs. To our knowledge, this has previously only been shown to be true for perceptual grouping of syllables. This suggests that the ITL can facilitate the formation of memories also for complex nonlinguistic stimuli.

Experiment 3: Perceptual Grouping of Visual Sequences

Experiment 3 investigates whether rhythmic grouping patterns generalize to the visual modality. It has been proposed that the ITL is not modality specific. For example, the grouping preference for syllables alternating in pitch into trochees and for syllables alternating in duration into iambs has also been observed in Italian participants with visual stimuli that alternate in temporal frequency and in duration (Peña, Bion, & Nespor, 2011). Even though there is no direct way to transfer the concept of pitch/fundamental frequency from acoustic to visual stimuli, humans are sensitive to the duration of visual stimuli and can track changes in temporal frequency in a sequence of visual stimuli (Kelly & Burbeck, 1984; Spillmann, 2006). Following Peña et al. (2011), we define *duration* as the time elapsed from the onset to the offset of a visual event (Lamb, 1991), and *temporal frequency* as the rate at which a visual stimulus changes its location on the screen (Hess & Snowden, 1992). Experiment 3 thus tests whether the grouping preferences of elements alternating in prominence defined by the ITL emerge also with visual stimuli cross-linguistically. We were particularly interested to see whether the grouping preferences for visual stimuli would be similar to the native listening effects we found in Experiment 1 with syllables or to the lack of native listening effects we found in Experiment 2 sine-wave analogues of speech. If the linguistic background does not influence the grouping of visual stimuli alternating in prominence, we would expect Italian- and Persian-speaking participants to group visual stimuli according to the ITL. This would support the hypothesis that the ITL is a domain general mechanism that is not influenced by domain-specific variation in acquired linguistic knowledge. Alternatively, it is also possible that the linguistic background of participants could influence participants' grouping preference if their native language has trochaic phrasal stress, as it influences their grouping of syllabic sequences. In such a scenario we would expect Persian- and Turkish-speaking participants to group visual events trochaically regardless of prominence type (temporal frequency or duration).

Method

Participants. We recruited further 24 Italian-speaking adults from Trieste (Italy), 24 Turkish-speaking adults from Istanbul (Turkey) and 24 Persian-speaking adults from Tehran (Iran), (23 females, mean age 22 ± 4 years). All participants were monolingual and reported no auditory, visual, or language related problems. None of the participants in Experiment 3 had taken part in either Experiment 1 or Experiment 2.

Stimuli. We created the visual sequence for the familiarization phase by concatenating 10 visual events and repeating them in the same order for 20 times. Each visual event was composed of an arrangement of 15 identical exemplars of a particular shape (e.g., 15 stars or 15 squares) randomly changing their location on the computer screen. There were 10 different shapes, one for each event. The sequence of shapes in the visual events consisted of a square, a four-pointed star, a cross, an arc, a triangle, a circle, a pentagon, a five-pointed star, a rectangle, and a diamond. This sequence was identical for all participants in all conditions. All visual events were presented in silver color against a black background. In the frequency condition, all visual events were presented for 800 ms (constant duration), with a brightness of 60–80 μW/cm2, but the shapes within each of the 10 visual events changed their location on the computer screen at either a high (25 Hz) or a low (10 Hz) temporal frequency (alternation in pitch). The high temporal frequency was attained by randomly changing the location of the shapes every 40 ms (25 Hz), whereas the low rate was achieved by changing the position of the shapes every 100 ms (10 Hz; see Figure 3). For the duration condition we used exactly the same sequences as in the frequency condition, however, the visual events alternated in duration (i.e., 320 or 800 ms), whereas temporal frequency remained constant (25 Hz and 60–80 W/cm2, respectively; see Figure 3). To ensure that the visual events were distinguishable in duration, we set the individual events at either 320 ms or 800 ms. Because the adjacent visual events alternated in duration in the duration condition and in visual frequency in the pitch condition, every switch in visual event (and in shape) matched a change in duration/frequency. The sequences were counterbalanced between participants (short event first/long event first) and faded in and out by scaling up/down the brightness of the videos during a 5-s period at the beginning and at the end.

For the test phase, we prepared a series of 10 static images composed of different pairings of the 10 shapes presented during familiarization. For each static image, one of the shapes of the pair was presented on the left side of the computer screen, and the other shape was presented on the right side of the screen (see Figure 3). We assumed that when asked to judge their temporal order, Turkish and Italian participants would interpret the image on the left of the screen as temporally preceding the image on the right (Boroditsky, 2001), but Iranian participants would interpret the images on the right of the screen as temporally preceding the image on the left. The 10 combinations of shapes were subdivided into two types of test items: (a) five adjacent test items consisting of pairs of shapes that appeared in adjacent visual events during familiarization and followed the iambic grouping principle; (b) five adjacent test items consisting of pairs of shapes that appeared adjacently during familiarization and followed the trochaic grouping principle. Even though all the test pairs had occurred adjacently during the familiarization, in the trochaic pairs in the frequency condition straddled the iambic pairs in the duration condition. The test images were static and did not contain any information about temporal frequency or duration

Procedure. Participants were tested in a silent room. Half of the participants from each language group were assigned to the frequency condition and half to the duration condition. Participants were instructed to watch carefully because they would be queried about the shape pairs they saw during the familiarization phase. Visual sequences were presented in the center (in a 270 × 230-mm2 area) of a 21-inch (53.34-cm), 1,024 × 768-pixel resolution screen. Participants were seated at a distance of 60–70 cm from the screen, allowing the visual events to be presented in the fovea (i.e., visual events subtended between 1.6° and 1.9° of visual angle in height and between 1.5° and 1.6° in width). After watching the familiarization stream, participants were presented with the test stimuli consisting of one pair of pictures at a time. Participants had to indicate yes/no to whether they had seen the pair of shapes in the given order during familiarization. Because all 10 shape pairs presented during test were adjacent during the familiarization—but half of the test shape pairs (half of the trials) followed iambic and the other half trochaic grouping—any consistent choice would be extremely strong evidence for perceptual grouping of visual stimuli. Each test item was presented twice and the order of the test items was randomized across participants. The experiment was controlled by PsyScope X software (http://psy.ck.sissa.it/).

Results and Discussion

The results of Experiment 3 are shown in Figure 4. The procedure for analyzing the results was identical to that of Experiment 1 and 2. The results revealed that visual events alternating in frequency or duration are grouped according to the ITL regardless of participants' native language. Model selection revealed a significant main effect of discrimination between visual test event pairs adhering/not adhering to the ITL, $\chi^2(df = 6) = 19.40$, $p = .004$, but no interaction between the test events adherence to the ITL and participants' native language, $\chi^2(df = 6) = 3.74$, $p = .712$, or experimental condition, $\chi^2(df = 5) = 3.53$, $p = .619$. Participants grouped visual events alternating in frequency into

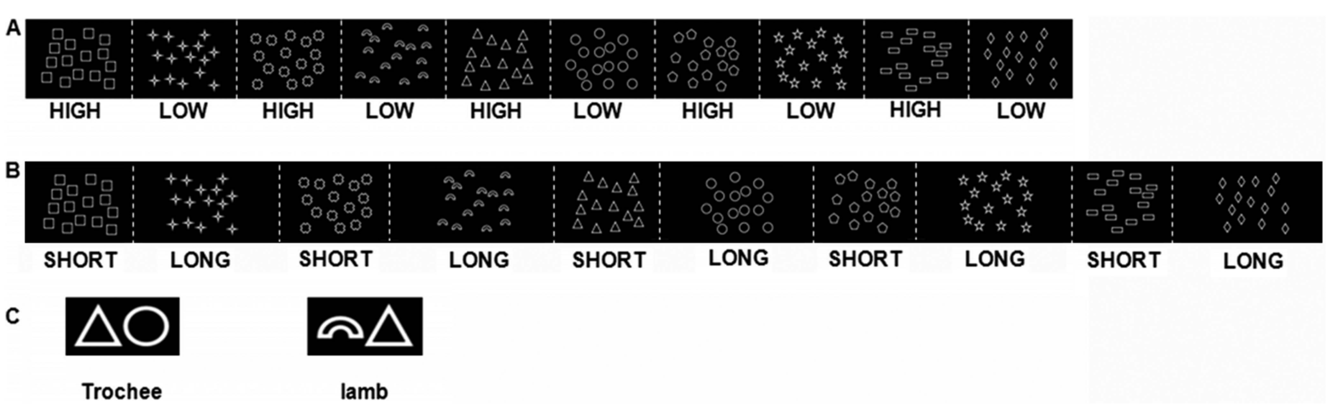

Figure 3. Stimuli of Experiment 3. The figure shows the fixed order of visual events within the visual sequences used in both conditions. In the pitch condition the visual events alternated in temporal frequency (A); in the duration condition the visual events alternated in duration (B). Figure 1C shows an example of the iambic/trochaic test items. Figure adapted from Peña, Bion, and Nespor (2011).

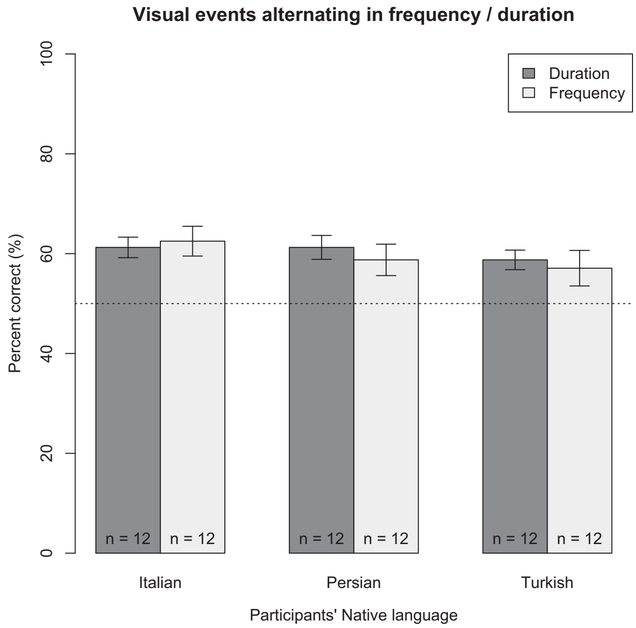

Figure 4. Italian- and Persian-speaking participants' grouping preferences of visual event sequences alternating in fundamental frequency and in duration. Y-axis denotes the percent of Iambic-Trochaic grouping: (a) correctly accepting high–low (strong–weak) pairs of visual shapes and rejecting low–high (weak–strong) pairs of visual shapes in the fundamental frequency condition; and (b) correctly accepting short–long (weak–strong) and rejecting long–short (strong–weak) pairs of visual shapes in the duration condition. Error bars denote standard error. The asterisk (*) denotes $p < .05$ against chance.

trochees regardless of their native language. We found a significant main effect of discrimination between test events adhering/not adhering to the ITL, $\chi^2(df = 1) = 14.43$, $p < .001$, but no interaction between test events adherence to the ITL and participants' native language, $\chi^2(df = 2) = 1.45$, $p = .486$. Italian-speaking participants accepted trochaic (high-low) and rejected iambic (low-high) visual events alternating in frequency significantly above chance ($d' = 1.43$, $SE = .45$, $z = 3.16$, $p < .005$) just like Persian- ($d' = 1.03$, $SE = .43$, $z = 2.40$, $p = .016$) and Turkish-speaking adults ($d' = 1.28$, $SE = .44$, $z = 2.90$, $p < .005$). Participants grouped visual events alternating in duration into iambs regardless of their native language. We found a significant main effect of discrimination between visual events adhering/not adhering to the ITL, $\chi^2(df = 1) = 5.89$, $p = .015$, but no significant interaction between test events' adherence to ITL and participants native language, $\chi^2(df = 2) = 0.046$, $p = .977$. Italian-speaking participants accepted iambic (short-long) and rejected trochaic (long-short) visual event pairs significantly above chance ($d' = 0.87$, $SE = .35$, $z = 2.46$, $p = .014$), as did Persian- ($d' = 0.89$, $SE = .35$, $z = 2.56$, $p = .0124$) and Turkish-speaking adults ($d' = 0.89$, $SE = .35$, $z = 2.48$, $p = .012$).

The results of Experiment 3 support previous findings with Italian-speaking adults who grouped visual events alternating in temporal frequency as strong-weak (trochaic grouping) and visual events alternating in duration as weak-strong (iambic grouping) (Peña et al., 2011). Also Turkish- and Persian-speaking adults grouped visual stimuli according to the ITL. This may suggest that the grouping of visual events alternating in prominence according to the ITL is a universal characteristic of visual perception. Importantly, because Italian-, Turkish-, and Persian-speaking adults grouping preferences of visual events parallel those we found for sine-wave speech in Experiment 2, the ITL is a domain general perceptual grouping principle that is independent of the modality (nonlinguistic auditory or visual) in which the rhythmic alternation is perceived. Because participants grouped visual events alternating in temporal frequency (Experiment 3) as they grouped auditory stimuli alternating in fundamental frequency (Experiment 1 and 2), our results also corroborate the idea that for perceptual grouping of elements alternating rhythmically, the temporal frequency of visual stimuli may be one correlate of fundamental frequency in auditory stimuli (Peña et al., 2011).

General Discussion

The results of Experiment 1 demonstrate that the prosody of spoken language can cause native language effects at the suprasegmental level. Despite being familiarized with exactly the same syllable sequences that alternated either in pitch or in duration, the grouping of Italian participants differed from that of both Turkish and Persian participants. Italian-speaking participants grouped syllables into pairs according to the ITL: Italian speakers grouped syllable pairs alternating in pitch into trochees (strong–weak/high–low) and syllable pairs alternating in duration into iambs (weak–strong/short–long). However, although Turkish and Persian-speaking adults grouped syllables alternating in pitch into trochees, they violated the ITL by grouping into trochees also syllables alternating in duration. Turkish and Persian signal prominence trochaically only within phonological phrases, where it is manifested through pitch and intensity on the syllable of the first word that bears lexical stress (Nespor et al., 2008; Nespor & Vogel, 2007; Windfuhr, 1979; cf. Table 1). Turkish- and Persian-speaking participants must therefore have relied on the trochaic rhythm of their native language also when grouping syllables alternating in duration. Our results, which clearly show that Turkish and Persian-speaking adults group nonnative linguistic rhythmic cues natively, are the first clear evidence for native listening effects in perceiving linguistic rhythm.

Native listening effects in perceiving prosody could interact with language acquisition in a number of possible ways. Because there are no attested violations in perceiving trochees signaled through pitch or intensity, the phrasal prosody of OV languages such as Japanese and Turkish, where phonological phrase prominence is primarily signaled through pitch and intensity, should be perceived in the same way independently of the participants' native language. Instead, the perception of VO languages such as English and Italian, where phrasal prominence is primarily signaled through duration that is perceived differently by speakers of OV and VO language (Nespor et al., 2008), could depend on the linguistic background of the listener. While the speakers of VO languages should not have difficulties with the phrasal prosody of OV languages, speakers of OV languages could demonstrate difficulties in correctly perceiving the phonological phrase prominence in VO languages. Because phrasal prosody provides cues to the hierarchical structure of syntax (cf. Nespor & Vogel, 2007; Langus et al., 2012), the native listening effects we observe in phrasal prominence could therefore influence how adult speakers acquire iambic and trochaic languages.

In contrast to the robust native listening effects with linguistic stimuli, the results of Experiment 2 and 3 demonstrated that the participants' native language does not influence the grouping of sine-wave speech and visual events. Instead, Italian-, Turkish-, and Persian-speaking adults grouped nonlinguistic stimuli according to the ITL. This contrasts with previous findings with speakers of Japanese—also a language that has trochaic phonological phrase prominence—who failed to group tones alternating in duration (Kusumoto & Moreton, 1997; Iversen, Patel, & Ohgushi, 2008). Because the trochaic grouping preference we found in Turkish- and Persian-speaking adults did not cause a trochaic preference for sine-wave speech and visual events alternating in duration, our results rule out the possibility that the knowledge of native language rhythm automatically influences the rhythmic grouping of nonlinguistic sounds. If cross-cultural variation would be attested in the perception of visual shapes or nonlinguistic sounds, then it may stem from culture specific knowledge acquired in nonlinguistic auditory or visual domains. Because the phrasal prosody of Japanese is similar to that of Turkish and Persian, understanding why Japanese-speaking participants' fail to group tones alternating in duration, could clarify whether native listening is domain specific also in music.

This does not imply that language cannot influence our perception in nonlinguistic domains. The tritone paradox is clearly an example that shows that listeners with different linguistic backgrounds have categorically different perceptions of exactly the same tone combinations. Investigating whether the tritone paradox (Deutsch, 1986) influences also the perceptual grouping of elements alternating in pitch could be useful in determining at what level of perceptual processing rhythmic grouping occurs. For example, what would happen if listeners, whose linguistic background causes different perceptions of the tritone interval, listened to a sequence of tones that alternate in prominence by a half-octave (a tritone)? Would participants who actually perceive the opposite pattern—ascending or descending—group these tones in opposite pairs? Or is rhythmic grouping based on the actual prominence of the stimulus rather than the perceived one? We know of no experimental work that would favor either one of these outcomes. However, determining the stage of perceptual processing at which elements alternating in prominence are grouped into either iambs or trochees could clearly help to clarify how the acquired knowledge of our native language interacts with basic perceptual laws such as the ITL.

It has been argued that native listening emerges early in cognitive development because infants begin to acquire knowledge about the language they are exposed to at birth, if not before (Cutler, 2012). For example, experimental evidence suggests that young infants group syllables alternating in intensity (Hay & Saffran, 2012; Yoshida et al., 2010) or in pitch (Bion et al., 2011) according to the ITL by 7-months of age. But they fail to group syllables and tones alternating in duration at least until 9 months-of age (Bion et al., 2011; Hay & Saffran, 2012; Yoshida et al., 2010). Although it is unclear how English- and Italian-learning infants acquire the iambic grouping principle, the results show that they do not group syllables alternating in duration iambically, like Persian- and Turkish-speaking adults in Experiment 1. This suggests that if assimilation to acquired knowledge is evidence for native listening, then native listening effects—at least for duration—do not emerge in speech perception of linguistic rhythm until the last months of the first year of life. This fits well with the idea that during the first year of life young infants are still tuning into their native language—a window of opportunity during which they are sensitive to many of the distinctions in linguistic sounds, even though these are not found in the language they are acquiring (Werker & Tees, 1984; Maurer & Werker, 2013).

The comparison between human adults' and infants' perception of the ITL thus raises questions regarding the extent native listening influences our perception of acoustic stimuli. For example, elements alternating in pitch appear to be universally grouped into trochees regardless of either the nature of the elements, the linguistic and cultural background or possibly even the developmental stage of the listener. The trochaic grouping of elements alternating in pitch could be a single domain general perceptual principle. However, the absence of native listening effects does not necessarily imply that listening to elements alternating in pitch relies on a domain general perceptual mechanism. It is also possible that the trochaic bias is encoded separately in the domain of language, in the domain of music, in the domain of nonlinguistic and nonmusical perception, and in the domain of vision.

However, future work on the ITL may prove useful in determining to what extent we perceive things natively and to what extent we rely on domain general principles. For example, evidence for a trochaic bias with pitch in newborn human infants could show that it is not learned. Alternatively, one could also investigate to what extent the nonviolated aspects of ITL are domain general. We imagine that transfer tasks, where participants have to decide on the rhythmic grouping of elements in one domain based on examples from another, could be a possible starting point. Such tasks would help to determine whether the ITL is a single perceptual mechanism that is not sensitive to the domain of the input (e.g., language, tones, sine-waves), or whether it is a domain-specific mechanism that is sensitive to the domain of the input but happens to yield identical groupings of elements alternating prominence across perceptual domains. The gray area where the theories about native listening and domain generality overlap could thus yield valuable information about how the human brain is tailored to accommodate acquired knowledge alongside universal principles of perception.

Taken together, the experiments reported in this manuscript show that native listening effects do occur in the perception of rhythm. However, the knowledge of our native language appears only to influence the perception of rhythm over linguistic stimuli, but not over nonlinguistic acoustical or visual stimuli. Acquiring our native tongue does not automatically distort our perception in other domains even when we consider a perceptual mechanism that, as far as we know so far, appears to be domain general. Restricting the ability of listening natively across perceptual domains may be necessary to avoid the cross-linguistic variation from interfering with the perception of the world more generally.

References

Ayat Hosseini, S. (2014). *The phonology and phonetics of prosodic prominence in Persian*. PhD Thesis, University of Tokyo.

Bates, D. (2005). Fitting linear models in R: Using the lme4 package. *R News, 5,* 27–30.

Bent, T., Bradlow, A. R., & Wright, B. A. (2006). The influence of linguistic experience on the cognitive processing of pitch in speech and nonspeech sounds. *Journal of Experimental Psychology: Human Perception and Performance, 32,* 97–103.

Bhatara, A., Boll-Avetisyan, N., Agus, T., Höhle, B., & Nazzi, T. (2015). Language experience affects grouping of musical instrument sounds. *Cognitive Science*. Advance online publication. http://dx.doi.org/10.1111/cogs.12300

Bhatara, A., Boll-Avetisyan, N., Unger, A., Nazzi, T., & Höhle, B. (2013). Native language affects rhythmic grouping of speech. *The Journal of the Acoustical Society of America, 134,* 3828–3843. http://dx.doi.org/10.1121/1.4823848

Bion, R. A. H., Benavides-Varela, S., & Nespor, M. (2011). Acoustic markers of prominence influence infants' and adults' segmentation of speech sequences. *Language and Speech, 54,* 123–140. http://dx.doi.org/10.1177/0023830910388018

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International, 5,* 341–345.

Boersma, P., & Weenink, D.. (2009). Praat: Doing phonetics by computer (Version 5.1.1) [Computer program]. Retrieved from http://www.praat.org

Bolton, T. (1894). Rhythm. *The American Journal of Psychology, 6,* 145–238. http://dx.doi.org/10.2307/1410948

Boroditsky, L. (2001). Does language shape thought? Mandarin and English speakers' conceptions of time. *Cognitive Psychology, 43,* 1–22. http://dx.doi.org/10.1006/cogp.2001.0748

Cutler, A. (1990). Exploiting prosodic probabilities in speech segmentation. In G. Altmann (Ed.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives* (pp. 105–121). Cambridge, MA: MIT Press.

Cutler, A. (2012). *Native listening*. Cambridge, MA: MIT Press.

Cutler, A., & Carter, D. M. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech & Language, 2,* 133–142. http://dx.doi.org/10.1016/0885-2308(87)90004-0

Cutler, A., Mehler, J., Norris, D., & Segui, J. (1983). A language-specific comprehension strategy. *Nature, 304,* 159–160. http://dx.doi.org/10.1038/304159a0

Cutler, A., Mehler, J., Norris, D., & Segui, J. (1986). The syllable's differing role in the segmentation of French and English. *Journal of Memory and Language, 25,* 385–400. http://dx.doi.org/10.1016/0749-596X(86)90033-1

Cutler, A., & Norris, D. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance, 14,* 113–121. http://dx.doi.org/10.1037/0096-1523.14.1.113

DeCarlo, L. T. (1998). Signal detection theory and generalized linear models. *Psychological Methods, 3,* 186–205. http://dx.doi.org/10.1037/1082-989X.3.2.186

DeCarlo, L. T. (2010). On the statistical and theoretical basis of signal detection theory and extensions: Unequal variance, random coefficient and mixture models. *Journal of Mathematical Psychology, 54,* 304–313. http://dx.doi.org/10.1016/j.jmp.2010.01.001

Deutsch, D. (1986). An auditory paradox. *Journal of the Acoustical Society of America, 80,* s93.

Deutsch, D. (1991). The tritone paradox: An influence of language on music perception. *Music Perception, 8,* 335–347. http://dx.doi.org/10.2307/40285517

Deutsch, D. (2013). The processing of pitch combinations. In D. Deutsch (Ed.), *The psychology of music* (3rd ed., pp. 249–325). San Diego, CA: Elsevier. http://dx.doi.org/10.1016/B978-0-12-381460-9.00007-9

Deutsch, D., North, T., & Ray, L. (1990). The tritone paradox: Correlate with the listener's vocal range for speech. *Music Perception, 7,* 371–384. http://dx.doi.org/10.2307/40285473

Dupoux, E., Kakehi, K., Hirose, Y., Pallier, C., & Mehler, J. (1999). Epenthetic vowels in Japanese: A perceptual illusion? *Journal of Experimental Psychology: Human Perception and Performance, 25,* 1568–1578. http://dx.doi.org/10.1037/0096-1523.25.6.1568

Dupoux, E., Pallier, C., Sebastián, N., & Mehler, J. (1997). A destressing "deafness" in French? *Journal of Memory and Language, 36,* 406–421. http://dx.doi.org/10.1006/jmla.1996.2500

Dupoux, E., Peperkamp, S., & Sebastián, N. (2001). A robust method to study stress 'deafness'. *Journal of the Acoustical Society of America, 110,* 1606–1618.

Dupoux, E., Sebastián-Gallés, N., Navarrete, E., & Peperkamp, S. (2008). Persistent stress 'deafness': The case of French learners of Spanish. *Cognition, 106,* 682–706. http://dx.doi.org/10.1016/j.cognition.2007.04.001

Dutoit, T., Pagel, V., Pierret, N., Bataille, F., & Van Der Vreken, O. (1996). The MBROLA project: Towards a set of high-quality speech synthesizers free of use for non-commercial purposes. In *Proc. ICSLP'96* (Vol. 3, pp. 1393–1396). Philadelphia, PA.

Fraisse, P. (1956). *Les Structures Rhythmiques*. Louvain, France: Publication Universitaires de Louvain.

Gerken, L. (1994). A metrical template account of children's weak syllable omissions from multisyllabic words. *Journal of Child Language, 21,* 565–584. http://dx.doi.org/10.1017/S0305000900009466

Goto, H. (1971). Auditory perception by normal Japanese adults of the sounds "L" and "R". *Neuropsychologia, 9,* 317–323. http://dx.doi.org/10.1016/0028-3932(71)90027-3

Gussenhoven, C., & Chen, A. J. (2000). Universal and language-specific effects in the perception of question intonation. In *International Conference on the Processing of Spoken Language (ICSLP) 6* (vol. II, pp. 91–94). Beijing, China.

Hay, J. S., & Diehl, R. L. (2007). Perception of rhythmic grouping: Testing the iambic/trochaic law. *Perception & Psychophysics, 69,* 113–122. http://dx.doi.org/10.3758/BF03194458

Hay, J. F., & Saffran, J. R. (2012). Rhythmic grouping biases constrain infant statistical learning. *Infancy, 17,* 610–641. http://dx.doi.org/10.1111/j.1532-7078.2011.00110.x

Hayes, B. (1995). *Metrical stress theory: Principles and case studies*. Chicago, IL: University of Chicago Press.

Hess, R. F., & Snowden, R. J. (1992). Temporal properties of human visual filters: Number, shapes and spatial covariation. *Vision Research, 32,* 47–59. http://dx.doi.org/10.1016/0042-6989(92)90112-V

Iversen, J. R., Patel, A. D., & Ohgushi, K. (2008). Perception of rhythmic grouping depends on auditory experience. *The Journal of the Acoustical Society of America, 124,* 2263–2271. http://dx.doi.org/10.1121/1.2973189

Jones, M. R. (1981). A tutorial on some issues and methods in serial pattern research. *Perception & Psychophysics, 30,* 492–504. http://dx.doi.org/10.3758/BF03204846

Jusczyk, P. W., Cutler, A., & Redanz, N. J. (1993). Infants' preference for the predominant stress patterns of English words. *Child Development, 64,* 675–687. http://dx.doi.org/10.2307/1131210

Kabak, R., & Vogel, I. (2001). The phonological word and stress assignment in Turkish. *Phonology, 18,* 315–360. http://dx.doi.org/10.1017/S0952675701004201

Kahnemuyipour, A. (2003). Syntactic categories and Persian stress. *Natural Language and Linguistic Theory, 21,* 333–379. http://dx.doi.org/10.1023/A:1023330609827

Kelly, D. H., & Burbeck, C. A. (1984). Critical problems in spatial vision. *Critical Reviews in Biomedical Engineering, 10,* 125–177.

Kornfilt, J. (1997). *Turkish*. London, UK: Routlege.

Kusumoto, K., & Moreton, E. (1997). Native language determines the parsing of nonlinguistic rhythmic stimuli. *The Journal of the Acoustical Society of America, 102,* 3204. http://dx.doi.org/10.1121/1.420936

Lamb, T. D. (1991). The role of photoreceptors in light-adaptation and dark-adaptation of the visual system. In C. Blakemore (Ed.), *Vision coding and efficiency* (pp. 161–168). Cambridge, UK: Cambridge University. http://dx.doi.org/10.1017/CBO9780511626197.017

Langus, A., Marchetto, E., Bion, R. A. H., & Nespor, M. (2012). Can prosody be used to discover hierarchical structure in continuous speech? *Journal of Memory and Language, 66,* 285–306. http://dx.doi.org/10.1016/j.jml.2011.09.004

Lee, H., & Noppeney, U. (2011). Physical and perceptual factors shape the neural mechanisms that integrate audiovisual signals in speech comprehension. *The Journal of Neuroscience, 31,* 11338–11350. http://dx.doi.org/10.1523/JNEUROSCI.6510-10.2011

Makarova, V. (2001). Perceptual correlates of sentence type intonation in Russian and Japanese. *Journal of Phonetics, 29,* 137–154. http://dx.doi.org/10.1006/jpho.2001.0137

Maurer, D., & Werker, J. F. (2013). Perceptual narrowing during infancy: A comparison of language and faces. *Developmental Psychobiology, 56,* 154–178.

Möttönen, R., Calvert, G. A., Jääskeläinen, I. P., Matthews, P. M., Thesen, T., Tuomainen, J., & Sams, M. (2006). Perceiving identical sounds as speech or non-speech modulates activity in the left posterior superior temporal sulcus. *NeuroImage, 30,* 563–569. http://dx.doi.org/10.1016/j.neuroimage.2005.10.002

Nespor, M. (1993). *Fonologia*. Bologna, Italy: il Mulino.

Nespor, M., Shukla, M., van de Vijver, R., Avesani, C., Schraudolf, H., & Donati, C. (2008). Different phrasal prominence realizations in VO and OV languages. *Lingue e Linguaggio, 2,* 139–168.

Nespor, M., & Vogel, I. (2007). *Prosodic phonology*. Berlin, Germany: Mouton De Gruyter. (Original work published 1986.) http://dx.doi.org/10.1515/9783110977790

Ordin, M., & Nespor, M. (2013). Transition probabilities and different levels of prominence in segmentation. *Language Learning, 63,* 800–834. http://dx.doi.org/10.1111/lang.12024

Patel, A. D. (2003). Language, music, syntax and the brain. *Nature Neuroscience, 6,* 674–681. http://dx.doi.org/10.1038/nn1082

Patel, A. D. (2008). *Music, Language, and the Brain*. New York, NY: Oxford University Press.

Patel, A. D., Peretz, I., Tramo, M., & Labreque, R. (1998). Processing prosodic and musical patterns: A neuropsychological investigation. *Brain and Language, 61,* 123–144. http://dx.doi.org/10.1006/brln.1997.1862

Peña, M., Bion, R. A. H., & Nespor, M. (2011). How modality specific is the iambic–trochaic law? Evidence from vision. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37,* 1199–1208. http://dx.doi.org/10.1037/a0023944

Peperkamp, S., & Dupoux, E. (2002). A typological study of "stress deafness". In C. Gussenhoven & N. Warner (Eds.), *Laboratory Phonology 7* (pp. 203–240). Berlin, Germany: Mouton de Gruyter. http://dx.doi.org/10.1515/9783110197105.203

R Development Core Team. (2012). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/

Remez, R. E., Dubowski, K. R., Broder, R. S., Davids, M. L., Grossman, Y. S., Moskalenko, M., . . . Hasbun, S. M. (2011). Auditory-phonetic projection and lexical structure in the recognition of sine-wave words. *Journal of Experimental Psychology: Human Perception and Performance, 37,* 968–977. http://dx.doi.org/10.1037/a0020734

Remez, R. E., Rubin, P. E., Pisoni, D. B., & Carrell, T. D. (1981). Speech perception without traditional speech cues. *Science, 212,* 947–949. http://dx.doi.org/10.1126/science.7233191

Revithiadou, A., Kaili, H., Prokou, S., & Tiliopoulou, A. (2006). Turkish accentuation revisited: A compositional approach to Turkish stress. In S. Yağcioğlu, A. Cem Değer, Ö. Kosaner, & A. Celtek (Eds.), *Advances in Turkish Linguistics: Proceedings of the 12th International Conference on Turkish Linguistics* (pp. 37–50). Izmir, Turkey: Dokuz Eylül Yayinları.

Spillmann, L. (2006). From perceptive fields to Gestalt. *Progress in Brain Research, 155,* 67–92. http://dx.doi.org/10.1016/S0079-6123(06)55005-8

Tanner, W. P., Jr., & Swets, J. A. (1954). A decision-making theory of visual detection. *Psychological Review, 61,* 401–409. http://dx.doi.org/10.1037/h0058700

Toro, J. M., Pons, F., Bion, R. A. H., & Sebastián-Gallés, N. (2011). The contribution of language-specific knowledge in the selection of statistically-coherent word candidates. *Journal of Memory and Language, 64,* 171–180. http://dx.doi.org/10.1016/j.jml.2010.11.005

Werker, J. F., Gilbert, J. H. V., Humphrey, K., & Tees, R. C. (1981). Developmental aspects of cross-language speech perception. *Child Development, 52,* 349–355. http://dx.doi.org/10.2307/1129249

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior & Development, 7,* 49–63. http://dx.doi.org/10.1016/S0163-6383(84)80022-3

Windfuhr, G. L. (1979). Persian grammar: History and state of its study. *Trends in Linguistics 12*. The Hague: Mouton Publishers. http://dx.doi.org/10.1515/9783110800425

Wong, P. C., Ciocca, V., Chan, A. H., Ha, L. Y., Tan, L.-H., & Peretz, I. (2012). Effects of culture on musical pitch perception. *PLoS ONE, 7,* e33424. http://dx.doi.org/10.1371/journal.pone.0033424

Woodrow, H. (1951). Time perception. In S. S. Stevens (Ed.), *Handbook of experimental psychology* (pp. 1224–1236). New York, NY: Wiley.

Yoshida, K. A., Iversen, J. R., Patel, A. D., Nito, H., Mazuka, R., Gervain, J., & Werker, J. F. (2010). The development of perceptual grouping in infancy: A Japanese-English cross-linguistic study. *Cognition, 115,* 356–361. http://dx.doi.org/10.1016/j.cognition.2010.01.005

Received April 8, 2014
Revision received September 27, 2015
Accepted October 26, 2015 ■