Review

# The missing link in the embodiment of syntax: Prosody

Hamutal Kreiner [a,b,*], Zohar Eviatar [a]

[a] Institute of Information Processing and Decision Making, Haifa University, Israel
[b] Linguistic Cognition Lab, Ruppin Academic Center, Israel

## A R T I C L E   I N F O

## A B S T R A C T

Neurolinguistic theories are challenged by the amodal abstract representations assumed by linguists. Embodied linguistics offers a conceptualization of the relationship between linguistic representation, experience, and the brain. Findings correlating brain activation patterns with referential features of words (e.g., body parts), suggest that the mechanism underlying linguistic embodiment is an "action–perception simulation". This mechanism accounts for embodied representation of words, but is harder to adapt to syntactic abstractions. We suggest that prosody is the missing link. Prosody is a sensory-motor phenomenon that can evoke an "action–perception simulation" that underpins the syntax-experience-brain association. Our review discusses different embodiment models and then integrates psycholinguistic and neurocognitive studies into a new approach to linguistic embodiment. We propose a novel implementation of the syntax-experience-brain relationship via the mapping between the temporo-spectral aspects of speech prosody and temporo-spectral patterns of synchronized behavior of neural populations. We discuss the potential implications for psycho- and neuro-linguistic research.

© 2014 Elsevier Inc. All rights reserved.

## Contents

* Corresponding author. Address: Linguistic Cognition Lab, Ruppin Academic Center, 40250 Emeque Hefer, Israel.
   E-mail address: hamutalk@ruppin.ac.il (H. Kreiner).

## 1. Introduction

The view of linguistic representation is going through a dramatic change in recent years. Early amodal approaches assumed abstract symbolic representation (e.g., Kintsch & van Dijk, 1978). More recent views assume that linguistic knowledge is represented as statistical distributions of word frequencies or even frequencies of word strings (e.g., Burgess & Lund, 1997). Currently, one of the prominent views, the embodiment approach, assumes that linguistic representations are grounded in the brain's modal systems (e.g., Barsalou, 2008a; Prinz, 2002). The theoretical appeal of this view compared to other, amodal, views, is that it offers a simple conceptualization of the relationship between representation, experience and the brain. In this paper we focus on the representation of syntax.[1] We ask what aspect of the physical events that constitute the experience of language is the most closely associated with syntax, and how is it grounded in the brain's modal systems. We suggest that prosody, the rhythmic and melodic aspect of spoken language, is the perceived physical event most closely associated with syntax. Prosodic features such as stress, pitch, and rhythm organize the acoustic stream within temporal structures that distinguish between grammatical categories and group words into phrases in a hierarchical structure. We posit that brain activity that resonates to these acoustic features of prosody underpins the embodiment of syntax.

In introducing our proposal we will briefly present the grounded cognition approach and focus on the embodiment of language. We then review the evidence for the close relationship between syntax and prosody. Subsequently we will argue that the acoustic features of prosody, such as pitch, stress, and rhythm are physical signals that the brain can simulate using temporospectral patterns of synchronized behavior of neural populations. Hence, we will propose that the mapping of prosodic patterns by corresponding patterns of neural activation underpins syntactic representations in the brain.

## 2. Grounding language

Although one of the fundamental assumptions of current cognitive psychology is that cognitive phenomena, including mental representation, have neurophysiological correlates, it is not clear how the associations between brain functioning and experience elicit symbolic representations. It is precisely this gap that theories of embodied, or grounded, cognition seek to address. Such theories posit that symbolic representations are grounded in the physical world. For example, when an experience such as eating an apple occurs, brain states associated with the smell, taste, touch, eating movements and even internal aspects such as change in sugar level, are activated. The brain states associated with the different modalities of this experience are captured and integrated into a multimodal representation. This multi-modal representation or parts of it are reactivated when we think, remember or actually

encounter the object (apple) or action (eating) experienced earlier. Thus, linguistic grounding is viewed as the manner by which symbolic abstractions emerge from the association of experience and relevant brain activity. According to Barsalou (2008a), most accounts of grounded cognition focus on the essential role of simulation in grounding cognition, and "Simulation is the reenactment of perceptual, motor, and introspective states acquired during experience with the world, body, and mind" (Barsalou, 2008a p. 618). Thus, the underlying assumption here is that the association between brain states and motor and perceptual aspects of experience is an essential aspect of embodiment. The ability to simulate this association even in the absence of real input, by reactivating brain states that were captured and saved during experience, is the physical manifestation, or the embodiment, of abstract mental representation.

This logic has been used to explain findings in social cognition and emotion perception (e.g., Niedenthal, 2007). For example, the ability to comprehend the emotional state of another person is based on a simulation of the physical markers of the emotion, such as facial expression, body position, speech characteristics, etc. The use of this logic to conceptualize linguistic representations is critical in view of the claim that one of the fundamental and unique species-specific aspects of human language is the capacity to disengage from the here-and-now, and from the concrete aspects of objects and events (e.g., Nelson, 2003; Tomasello, 2003). The use of simulation allows linguistic knowledge, representation and even communication to be removed, or dissociated from the concrete perceivable context. It offers a solution to the impasse defined by the abstractness of language and the physical, measurable functions of the brain.

### 2.1. The embodiment of meaning vs. the grounding of structure

The discontinuity between the symbolic constituents used in cognitive models and the physical measures used to describe brain functioning is one of the reflections of the gap between mental representation and the physical world. The rise of the linguistic embodiment view, that offered a way to bridge this gap by assuming analogical rather than symbolic representations, has had an almost revolutionary impact on psycholinguistic research (for reviews see Barsalou, 2008a; Gibbs, 2003). This change, together with progress in neuroimaging methods, has driven the wealth of neurolinguistic imaging studies that explored how language is represented in the brain.

An overview of these studies suggests a distinction between two types of models that attempt to ground language onto patterns of brain activation in different ways. The majority of embodiment studies (reviewed in Barsalou, 2008a) examine the *embodiment of content*. Their major goal is to describe the content of semantic and lexical knowledge, and to do so they focus on the role of the body and the perceptual modalities in representing the meaning of words and expressions. These theories seem to neglect the basic regularities that characterize human languages, which are described by syntactic theories. Recently, however, several authors, such as Glenberg and Gallese (2012) and Pulvermüller (2010), offered a different kind of grounding. They address regularities such as the order of sentence constituents, their hierarchic relationships, and their dependencies, and therefore seem to focus on the *grounding of structure*. The two types of models share the

---

[1] Due to the large cross-linguistic diversity in the regularities considered as syntax (Evans & Levinson, 2009), and the ensuing disagreement between theories of syntax, it is hard to provide a very detailed definition of syntax. For the purpose of this article we refer to the very basic conception of syntax as the system underlying the universal human ability to compute nonlinear long range relationships between elements that occur linearly in time (e.g., segmentation, hierarchy and binding), assuming that different languages may use different systems of rules to achieve this goal.

assumption that language is grounded via simulation, or reactivation of brain states associated with motor, perceptual and internal experience. However, they seem to differ in their view of what is simulated. Models of *content embodiment* simulate the semantic aspects of language, anchoring our experience with objects and events to brain activity. Models of *structure grounding* simulate the structural aspects of language, anchoring the language-internal regularities to brain activity.

### 2.1.1. The embodiment of content and meaning

The *perceptual symbol system* is one of the fundamental mechanisms in Barsalou's view of embodied cognition. In this system, symbols are represented in an analogical fashion by the neural activity invoked by their perception and/or by associated action. Critically, simulation of such action–perception activity can be invoked in the absence of real action–perception, giving rise to *mental* representations (e.g., Barsalou, 2008a, 2008b). Thus, models of content embodiment focus on the role of the body in cognition. These models are supported by widespread findings that bodily states can both cause and be affected by mental representations (for reviews see Barsalou, 2008b; Glenberg, 2010; Willems & Hagoort, 2007). For example, Pulvermüller (2005) showed that when participants read an action-word, activation in the related motor system in the brain is increased such that verbs associated with head, arm, or leg actions invoked activation in the respective areas of the motor system. Such studies support grounding accounts that emphasize simulation of body states and situated action (e.g., Barsalou, 2003; Lakoff & Johnson, 2008; Niedenthal, 2007; Smith, 2005), and have recently inspired many insightful models of linguistic embodiment (Mahon & Caramazza, 2009; Pulvermüller & Fadiga, 2010; Zwaan, 2008).

One of the important bases of this view was the discovery of the mirror neuron system (MNS) and the proposal that it functions as a mediator between perception and action, underpinning mental representations including language. Thus, the MNS system is thought to be the fundamental mechanism underlying "action–perception simulation". Nevertheless, in discussing the relevance of MNS to the embodiment of syntax, Tettamanti and Moro (2012) note that due to the defining characteristics the MNS, it cannot subserve all linguistic functions. These authors argue that "the most important limitation is that some core defining properties of human language, at the phonetic, semantic, and especially at the syntactic level, are not transparent to the bodily senses and thus they cannot be the direct source of mirror neuron perceptuo-motor matching" (p. 923).

Another limitation of theories of semantic embodiment is that they neglect the basic regularities that characterize human languages, the regularities described by syntactic theories. Thus, language-internal regularities such as the systematic segmentation of utterances into constituents, as well as the order and hierarchical relationship between these constituents, are not addressed in such theories (but see Glenberg & Gallese, 2012; Pulvermüller, 2010). We suppose that the dearth of studies examining how syntactic abstractions emerge from brain mechanisms may result from the large gap between concrete brain mechanisms and the symbolic abstractions used to describe syntax. These abstractions largely ignore the physical aspects of language, namely its temporal dimension, as well as its acoustic realization and the mechanisms of production and perception of this acoustics.

### 2.1.2. The grounding of structural regularities

The models that we have classified as embodying structure rather than content take a different path and try to address some of the language-internal regularities. We focus on two recent models that addressed the embodiment of syntax: Pulvermüller's (2010) *Discrete Combinatorial Neuronal Assemblies* model that proposes a micro level mechanism and Glenberg and Gallese's (2012) *Action-based-language* model that proposes a macro level mechanism for the perception and production of speech.

#### 2.1.2.1. Discrete combinatorial neuronal assemblies.
Most, if not all, theoretical models of syntax do not attempt to implement the grammatical constructs they construe onto biological mechanisms, as they basically see them as amodal abstractions. Pulvermüller (2010) agrees with the view that syntax is an abstract system of rules, but argues that "its core principles can be translated into the profane language of nerve cells and circuits" (p. 167). Thus, he proposes that abstract combinatorial rules can be computed by patterns of activation in functionally discrete neuronal assemblies, and that this mechanism can account for the implementation of syntax in the brain. The "discrete combinatorial neuronal assemblies" (DCNAs) are composed of sequence detectors of the same type as motion sequence detectors in higher mammals. Reciprocal connections between these detectors and synchronized firing chains constitute the assemblies, and allow hierarchical links. Pulvermüller suggests that such assemblies offer an implementation of abstract principles such as recursion and embedding, which are fundamental characteristics of syntactic processing.

Pulvermüller's (2010) model is important because it offers a specific neural mechanism to underpin abstract principles that were previously considered as impossible to implement in a neuronal automaton (Chomsky, 1963). In that sense it may be considered as part of the paradigm of embodied linguistics. However, while this model offers direct implementation of abstract rules, it neglects the sensory-motor grounding that the embodiment approach offers to explain the association between experience and the brain. Thus, in Pulvermüller's model it is not clear what aspect of the actual input would trigger the use of the mechanism to represent the appropriate syntactic structure. For example, the utterance "*George ate the apples and the oranges*" may have the continuation "*were eaten by his sister*". In such garden-path sentences, the initial utterance by itself results in one structure, but the continuation results in a radical change to this structure. In sentences like this, order information may not be sufficient to support incremental construction of the correct hierarchical structure. Previous parsing models assumed either rule based (e.g., Minimal Attachment) or usage based (e.g., frequency) algorithms to describe the manner in which such potential ambiguities are resolved and hierarchical structures are construed. Alternatively, prosodic cues, such as pauses, which unfold over time, provide a perceptual input that serves as an early reliable and even predictive trigger to build one structure or the other, with minimal assumptions about top-down processes. However, the prosodic experience involved in such ambiguity resolution is not represented in Pulvermüller's model because it is basically a model that implements abstract syntactic knowledge without incorporating the experience that invokes it.

#### 2.1.2.2. Action-based-language model.
While Pulvermüller's (2010) model bridges the gap between representation and the brain by implementing abstract rules directly in neural mechanisms, it is not clear what role experience plays in this model. Glenberg and Gallese's (2012) *Action-Based-Language* model targets this gap and maps syntax to the brain via experience. They note that the traditional description of language in terms of "symbols (e.g., words) and rules for manipulating them (e.g., syntax)" easily explains some features of human language such as the fact that it is productive and compositional, but it has "a difficult time explaining meaning (how the abstract symbols are grounded)" (p. 919). Thus, they propose that action-based models offer a clear conceptualization of the relationship between meaning and experience. According to their model, "syntax emerges from modifying

hierarchical control of action to produce hierarchical control of speech."(p. 914). This model is consistent with the evolutionary view that neural circuits that evolved to regulate order and reiteration in motor control have an important role in speech production and syntax (e.g., Lieberman, 2007). However, although their model accounts for the acquisition and use of nouns and verbs, Glenberg and Gallese admit that a comprehensive account of syntactic knowledge "remains a promissory note" (Glenberg & Gallese, 2012, p. 914). It seems, then, that mechanisms of action control, complex though they are, are not enough to underpin the complex regularities conceptualized by syntax.

Both Pulvermüller's (2010) *Discrete Combinatorial Neuronal Assemblies* model and Glenberg and Gallese's (2012) *Action-based-language* model address linguistic regularities captured by syntax. However, they ignore the fundamental regularities of spoken language – the regularities of the acoustic aspects of language. Moreover, they ignore the close relationship between syntax and such acoustic regularities as stress, rhythm and intonation that constitute prosody. To understand how syntax is grounded, we need to specify the motor-sensory experience that can map onto the complex syntactic abstractions. We follow Glenberg and Gallese's (2012) view that the notion of action-based hierarchical organization is one of the key features in bridging the gap between brain mechanisms and syntax. Nevertheless, we believe that one of the unique characteristics of language, that distinguishes it from other aspects of action, is that it can represent information which is removed and dissociated from the present context, and syntax has a central role in this displacement. Hence, we argue that because syntax is mainly an organizational system and not a referential system (but see Goldberg, 2003) it is unlikely that it is grounded in the same way as the referential features of language (semantics, pragmatics). Rather, in what follows we propose that grounding of syntax is based on the association of the acoustic, and particularly prosodic, experience and the associated brain activity.

*2.1.2.3. Speech and brain oscillations.* We suggest that the work of Giraud and Poeppel (e.g., 2012) offers a new way of grounding linguistic regularities. Both language comprehension and production require simultaneous processing of hierarchical segments, with phonemes embedded in syllables, syllables in words, words in phrases, etc. In processing spoken language, hierarchical processing of the acoustic signal (or visual, in the case of sign languages, Kremers, 2012) requires representation and computation at different levels of temporal resolution. This simultaneous processing of acoustic segments at hierarchical time windows motivated different propositions by Poeppel et al. (e.g., Hickok & Poeppel, 2007) regarding implementation in anatomy and in brain mechanisms. In summarizing several works, Ghitza (2011) proposes that brain waves at different frequencies are correlated with different time windows that are related to different aspects of speech comprehension. Although these authors do not use the term *embodiment*, the neural mechanisms they propose associate the spectral patterns of oscillatory rhythms with the temporo-spectral aspects of speech, which are based on prosodic characteristics. Hence, their work offers a mechanism that associates speech prosody to brain activity. We propose that in view of the close relationship between speech prosody and syntax this association can be the substrate of syntactic representation.

In the next section we discuss the mapping between syntax and the acoustic regularities reflected in prosody. We then summarize recent findings that reveal the coordination between rhythms of speech and the periodicity of brain waves as measured by brain wave oscillations (e.g., Ghitza & Greenberg, 2009; Luo & Poeppel, 2012). Finally, we argue that syntax may be viewed as a representation emergent from activation and /or simulation of oscillatory rhythms associated with prosodic features of speech.

## 3. Syntax and prosody

The term prosody refers to the acoustic–phonetic properties of an utterance that "vary independently of its lexical items" (Wagner & Watson, 2010, pp. 905). The acoustic realization of each of the properties has a perceptual correlate, for example fundamental frequency is perceived as pitch, and intensity is perceived as loudness. Prosody generally refers to patterns of variation in these properties. For example, the term intonation refers to changes in pitch over a phrase, sentence or utterance, such as the pitch-rise typical of questions in English. More complex prosodic percepts, such as stress emphasis, pitch accenting, intonational breaks and rhythmic patterns are generated by synchronized co-variation of various acoustic properties. Thus prosody refers to the structural patterns of acoustic modulations over time (for an extended discussion of what constitutes prosody, see Wagner & Watson, 2010).[2] In this section we briefly review evidence for the close relationship between these modulations and syntactic regularities.

### 3.1. The role of prosody in the development of syntax

Even the most nativist approaches in psycholinguistics acknowledge that experience is crucial to language acquisition and that this experience is based essentially on auditory input (or visual input in the case of signed languages, see Sandler, Meir, Dachkovsky, Padden, & Aronoff, 2011). Prosody has been shown to have a major role in this process from the very beginning. Newborns as young as 3 days of age show a preference for the sound of their native language over foreign languages, and this was attributed to the familiarity of the prosodic patterns of their native language (Christophe, Mehler, & Sebastian-Galles, 2001; Mehler et al., 1988). These findings clearly indicate that prosodic sensitivity is exhibited very early and may serve as a precursor to language acquisition (Mehler et al., 1988).

More important, however, are the findings that indicate that infants use prosodic cues as signals of different linguistic properties. For example, Johnson and Jusczyk (2001) show that 8-month old infants use stress patterns as a cue for segmenting the speech stream into words. Mandel, Jusczyk, and Nelson (1994) show that 2-months-old not only prefer to listen to coherent prosodic phrases, but their memory for words presented in such prosody is better than for words presented in a list intonation. Finally, there is evidence to suggest that 9-month-olds prefer passages where the prosodic break corresponds to the major syntactic break compared to passages where prosodic and syntactic breaks mismatch (e.g., Jusczyk et al., 1992). The accumulating evidence from infants' pre-linguistic use of prosodic cues indicates that infants use such cues as an organizational device that helps them in segmenting the continuous speech stream and attracts their attention to different linguistic properties (for a recent review see Speer & Ito, 2009). Hence, Speer and Ito suggest that children use prosodic phrasal structure as a 'proto-syntax' that serves as the basis for later stages in language acquisition. Moreover, the important contribution of prosodic cues to language acquisition was recently demonstrated in adults as well. Langus, Marchetto, Bion, and Nespor (2012) show that in foreign language acquisition, adults can use prosodic cues such as pitch-declination and lengthening of final syllable, to learn

---

[2] Prosodic patterns may also reflect non-linguistic processes involved in speech such as breathing and emotional state, but here we focus on language related prosodic phenomena.

the regularities of a new language as reflected in its transitional probabilities.

### 3.2. Correspondence of prosodic and syntactic structures in language production

Studies of language processing show that prosody is not only important for acquisition, but that it also plays a central role for adults both in comprehension and in production of spoken language (for reviews see Cutler, Dahan, & van Donselaar, 1997; Speer & Blodgett, 2006; Wagner & Watson, 2010). In production, most of the early research focused on the correspondence between prosodic boundaries and syntactic structure. The evidence shows that the number and positions of pauses in an utterance are roughly correlated with the number and positions of syntactic phrase boundaries (Cooper & Paccia-Cooper, 1980; Nespor & Vogel, 1986; Selkirk, 1984), and that the durations of pauses is correlated with the strength of the syntactic boundaries (Grosjean, Grosjean, & Lane, 1979). Ladd (2008) reviews more recent findings showing that speakers produce a variety of different prosodic features that can be associated with the boundaries and hierarchic relationship of syntactic constituents; for example, lengthening of the pre-boundary syllable (e.g., Byrd, Krivokapic, & Lee, 2006) and a "reset" of intonation to its base pitch (de Pijper & Sanderman, 1994; Truckenbrodt, 2002).[3]

### 3.3. The contribution of prosody to syntactic processing in language comprehension

In comprehension, it has been suggested that prosody helps the listener by transforming the serial auditory input into structured patterns (Cutler et al., 1997) that facilitate the segmentation of the speech stream into words (e.g., Cutler, Butterfield, & Williams, 1987). Lexical prosody can facilitate the classification of words to grammatical categories (Kelly, 1992), for example different stress patterns characterize verbs and nouns like in the case of the verb *contrast* vs. the noun *contrast*. Most important to the current paper, is the finding that phrase-level prosody interacts with syntactic parsing in comprehension. Thus, comprehension is facilitated when the prosodic and the syntactic structures correspond, but when they conflict, comprehension is hindered (e.g., Carroll & Slowiaczek, 1987). Moreover, prosodic structure can bias or even determine the way syntactic ambiguity is resolved (e.g., Kjelgaard & Speer, 1999; Price, Ostendorf, Shattuck-Hufnagle, & Fong, 1991). For example, in sentences such as "*the boy touched the girl with the flower*" prosodic cues clarify whether *the flower* should be attached to *the girl* (*i.e., the girl with the flower* was touched) or to the verb *touched* (*i.e., the girl* was *touched with the flower*) (e.g., Pynte, 1996; Snedeker & Casserly, 2010). In other cases, later occurring context may disambiguate such syntactic ambiguities, as we demonstrated earlier with the sentence *George ate the apples and the oranges were eaten by his sister*". However, prosodic cues contribute to online disambiguation and to prediction of syntactic structure, allowing incremental processing as the utterance is unfolding, without waiting for the disambiguating context.

The contribution of prosody to comprehension has been explained in terms of different cognitive processes. Some accounts (e.g., Carlson, Frazier, & Clifton, 2009) focused on prosodic chunking and argued that certain prosodic features (e.g. slowing,

pausing) are involved in grouping units and consequently they constrain the parsing possibilities. Prosodic grouping may constrain the location and duration of syntactic phrase boundaries and the hierarchical relationship between syntactic constituents. Speer and colleagues (Speer & Ito, 2009; Speer et al., 2011) further suggested that prosodic chunking allows more efficient use of limited resources in working memory (e.g. Speer, Crowder, & Thomas, 1993).

Other accounts focused on prosodic prominence, conveyed by cues such as pitch and loudness, and showed that it may affect the resolution of binding relationships. For example, in the sentence *Sara called Gillian and then she came to visit*, the pronoun *she* can refer either to *Sara* or to *Gillian*, but listeners tend to bind it to the more prominent name, and prominence is often conveyed by prosodic cues. Thus, prosodic prominence is involved in focusing attention on a particular constituent facilitating its maintenance in working memory, and making it more accessible when the referring expression (e.g. *she*) is encountered. The contribution of prosodic prominence has been shown in binding processes such as pronoun resolution (e.g., Hirschberg & Ward, 1991) and in other cases of reference resolution (e.g., Dahan, Tanenhaus, & Chambers, 2002).

Finally, some accounts consider the prosodic structure as a whole, not focusing on particular aspects. For example, Aylett and Turk (2004) suggested that prosody enhances local predictability and thereby facilitates processing. It seems then that the rich and complex array of prosodic cues contribute not merely to signal different aspects of syntactic structure, but also to facilitate cognitive processes involved in language processing such as memory and attention. Frazier, Carlson, and Clifton (2006) suggested that the prosodic representation serves as a multi-layered map indexing items across different representations (phonological, syntactic, semantic), while the utterance is being processed. Kotz and Schwartze (2010) add that prosody captures the regularities of the acoustic signal that organize information within temporal structures. It seems then, that prosody is a perceptual signal that integrates cues at multiple temporal scales in a hierarchical organization that facilitates syntactic processing.

### 3.4. Implicit prosody and syntax

Recent findings indicate that the syntax-prosody relationship is not necessarily aimed solely at facilitating oral communication (e.g., Hwang & Steinhauer, 2011; Speer et al., 2011), and suggest that implicit prosody is also part of linguistic representation (Fodor, 2002). Consistent with this view, several theoretical and computational models consider the mapping between prosodic and syntactic representations as an *inherent* characteristic of the human language faculty, rather than an epiphenomenon of oral communication. These models differ in their assumptions about the relative contributions of lower level factors such as syllable structure and lexical prosody (e.g., Nespor & Vogel, 1986; Selkirk, 1984) and higher level factors such as semantics and pragmatics (e.g., Ferreira, 1993) in modulating prosody. Some models focus on mapping prosodic prominence (e.g., Chomsky & Halle, 1968), while others focus on phrase boundaries (e.g., Nespor & Vogel, 1986). Consequently, the models differ on how closely prosodic and syntactic structures are mapped onto each other. Interestingly, one of the most comprehensive models assumes that "intonation structure and surface structure are simply different aspects of the same derivational structure" (Steedman, 2000, p. 680).

The view that the associated syntax-prosody representations are not merely an epiphenomenon of oral communication is echoed in findings from silent reading studies. Evidence from such studies indicate that even when explicit prosody is not available (except for punctuation), "implicit prosody" influences syntactic

---

[3] It is not argued that speech prosody is driven exclusively by syntax. It has been shown that the prosodic patterns produced by speakers cannot be predicted from syntactic structure per se, and other linguistic and non-linguistic factors modulate the production of prosody (e.g., Ferreira, 1991; Speer, Warren, & Schafer, 2011; Watson & Gibson, 2004; for a review see Wagner & Watson, 2010).

parsing (Fodor, 2002). For example, Bader (1998) showed that when reading silently syntactically ambiguous sentences, readers have more difficulty when both syntax and prosody must be re-analyzed to make sense of the sentence (e.g., *Since Jay always jogs a mile and a half seems like a short distance to him*), than when the prosodic structure remains the same and syntax alone must be re-analyzed (and see also Steinhauer & Friederici, 2001). Fodor has further argued that implicit prosodic representations influence syntactic parsing decisions in a way that explains why speakers of different languages have different preferences in resolving syntactic ambiguities (Fodor, 2002; Hwang & Steinhauer, 2011).

Recent neuroimaging studies indicating that prosodic manipulations in silent reading activate voice selective areas in the auditory cortex (Yao, Belin, & Scheepers, 2011), are consistent with the idea that prosody is activated and represented in the brain even in silent reading, and not only when it is conveyed in the message as part of oral communication. In fact, Yao et al. (2011) suggest that such "spontaneous auditory imagery" may be interpreted as "perceptual simulation" in line with the embodied cognition approach.

### 3.5. Prosody and syntax in the brain

Typically, brain-syntax models focus on localization (e.g., Friederici, 2011; Grodzinsky, 2006, 2010), and there have been only few attempts to explore the mechanisms of syntactic processing in the brain (Pulvermüller, 2010; Pulvermüller & Fadiga, 2010). Nevertheless, there are neuropsychological studies that investigated the functional relationship between syntax and prosody in the brain. The robust evidence that syntax is lateralized to the LH combined with the accumulating evidence that linguistic prosody is lateralized to the RH are puzzling in view of the psycholinguistic evidence for the close relationship between prosody and syntax. This puzzle motivated the investigation of the role of the Corpus Callosum (CC) in the inter-hemispheric information exchange required for coordinating the lateralized speech streams.

Several case studies suggested that the CC is involved in inter-hemispheric exchange of auditory, and particularly prosodic information (Klouda, Robin, Graff-Radford, & Copper, 1988; Pollmann, Maertens, von Cramon, Lepsien, & Hughdahl, 2002). In a more systematic study Friederici, von Cramon, and Kotz (2007) used ERP measures and compared control participants and patients with lesions in the anterior or posterior portions of the CC to investigate the role of the CC in the interaction between syntax and prosody. The findings indicated that controls showed an N400-like effect when syntax violated the prosody-induced syntactic expectations but patients with lesions in the posterior CC did not show this effect. Critically, these patients demonstrated a prosody-independent semantic N400 effect. Hence the findings highlight the role of the posterior CC as the neuroanatomical structure for the inter-hemispheric interface between prosody and syntax.

Friederici et al. (2007) demonstrated the influence of prosodic information on syntactic processing. However, the prosody–syntax interface seems to be bi-directional such that prosody influences syntactic processing but also syntax influences prosodic processing, for example by predicting prosodic phrasing (for review, see Cutler et al., 1997; Eckstein & Friederici, 2006). Sammler, Kotz, Eckstein, Ott, and Friederici (2010) extended Friederici et al.'s (2007) investigation to examine whether the CC is involved in the reversed information exchange, namely the influence of syntax on prosody. As a marker for syntax–prosody interaction they used an anterior negativity elicited by a mismatch between syntactically predicted phrase structure and prosody. Their findings are similar to those of Friederici et al. (2007), while controls showed the anterior negativity, patients with posterior CC lesions failed to show this effect, although they showed intact, prosody-independent

syntactic processing comparable to controls. Sammler et al. (2010) concluded that homologous regions in the two hemispheres process syntax (left) and prosody (right) and that their activity is coordinated and integrated via the corpus callosum.

This conclusion was further supported by a study that used Dynamic Causal Modeling (DCM), a method of analyzing neuroimaging data to identify activation coupling among brain regions and to estimate to what extent this coupling is influenced by experimental changes. The findings from this analysis revealed connectivity between syntactic and prosodic processing, and were novel in suggesting that a deep subcortical structure, possibly thalamo-cortical (HG) loops, may be involved in interfacing prosody and syntax (David, Maess, Eckstein, & Friederici, 2011). Interestingly, the findings further suggest that when one hemisphere detects a violation, either syntactic or prosodic, the inter-callosal connectivity is reduced to allow independent processing of syntax (left hemisphere) and of prosody (right hemisphere).

## 4. The Embodiment of syntax via simulation of prosodic patterns

Given that there are functional connections between syntax and prosody in the brain, it may be easier to anchor syntactic processes in the brain via the temporo-spectral structure of prosody than without it. Hence in this last section we propose that syntax may be grounded in the brain via the simulation of prosodic patterns, we review some evidence associating brain activation to prosody in ways that may underpin the basic regularities represented by syntax. We begin our discussion with one of the earliest models for a mechanism that maps between abstract linguistic units and the sensory-motor system, namely Liberman's *Motor Theory of Speech Perception* (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967). Liberman et al. did not use the term embodiment, and their theory is not a theory of embodiment in the classical sense, associating patterns of brain activation and experience. Nevertheless, their theory may be viewed as an early version of embodiment because essentially, it proposed that perception of speech involves simulation of the motoric aspects of speech.

We examine the implications of this theory to the representation of language, and posit that it shares two important principles with more recent approaches to embodiment: The first is the notion of simulation, and the second is the idea that simulation can serve to implement abstract representations. Unlike the recent approaches to linguistic embodiment reviewed above, this theory is focused on simulation of the fundamental regularities that characterize the physical aspect of language – its acoustic realization. We attempt to transfer these principles to the grounding of syntax, and propose that an association between patterns of brain activation and the physical characteristics of prosodic patterns may allow an experience-based simulation from which abstractions of the fundamental regularities of language emerge.

### 4.1. The Motor Theory of Speech Perception as an embodiment model

Harris (1953) observed that speech cannot be synthesized from discrete phone-sized segments grouped like building blocks. The *Motor Theory of Speech Perception* (Liberman & Whalen, 2000; Liberman et al., 1967; for recent review see Galantucci, Fowler, & Turvey, 2006) was developed to address this issue, and aimed to answer the elementary questions about speech perception, namely, how is the continuous acoustic stream of speech segmented into linguistic units?

Liberman suggested that serial linear "segments" of speech are not an adequate input for speech perception because they vary according to the context in which they occur, like in the example

of the English phoneme /n/ that sounds differently in the word *tan* and in the word *tank*. This context-dependent phonetic variation is attributed to several processes, but the most frequently and universally mentioned process is coarticulation. Coarticulation occurs in continuous speech when successive consonants and vowels are articulated in partial overlap in time, as a result of the continuous nature of vocal tract gestures. Due to this context-dependent phonetic variation, listeners cannot recover discrete phone-sized segments that can be processed like a discrete "alphabet" from the continuous acoustic signal of the speech stream. According to the *Motor Theory of Speech Perception*, to enable such decoding, the vocal motor system that produces coarticulation is recruited. Liberman and Whalen (2000) described it as if "The gestures [of the vocal tract] are combined in various ways to form the phonetic segments, as if the gestures were the atoms of language and the segments its molecules"(p. 188). Thus phonetic segmentation is achieved through recovering the atomic gestures that combine to create the continuous acoustic speech stream.

### 4.1.1. Neural activity simulates experience

According to the *Motor Theory of Speech Perception,* in the perception of speech, gestures are "recovered from the acoustic signal" (Liberman & Whalen, 2000, pp. 189). While Liberman and his colleagues do not commit to how this recovery is implemented, one way to think about it is in terms of reenactment, or simulation of the gestures that generated the acoustic signal. Evidence from different neuro-imaging methods suggests that this may be implemented by activating neural mechanisms associated with speech related motor systems. For example, evidence from Functional Magnetic Resonance Imaging (fMRI) studies indicate that there is overlap between the cortical areas active during speech production and those active during passive listening to speech (e.g., Pulvermüller et al., 2006; Wilson, Saygin, Sereno, & Iacoboni, 2004). In reviewing the accumulating neuro-imaging evidence that support the idea of simulation of vocal tract gestures during speech perception, Galantucci et al. (2006) argue that Liberman's theory "anticipated a theme that has become widespread in cognitive science" (pp. 361).

### 4.1.2. Simulation opens the door to abstract representations

Mattingly and Liberman (1988) suggest that the human ability to recover the vocal gestures from the acoustic signal establishes the parity between listener and speaker and serves as a "common currency" (p. 190) that allows them to communicate. In discussing the notion of parity and its contribution to the development of the *Motor Theory of Speech Perception*, Galantucci et al. (2006) propose that "parity is intended to be an abstract constraint on the symmetric co-evolution of the machinery for producing and perceiving speech". Thus it seems that abstraction is implemented by the simulation of gestures recovered from the acoustic signal, like in recent models of linguistic embodiment. Moreover, it seems that the ability to invoke such simulations serves as a means of sharing and communicating with others; a function similar to that proposed by embodiment models in social cognition.

### 4.1.3. Embodiment of linguistic regularities via simulation of sound patterns

The insights from the *Motor Theory of Speech Perception* paved the way for the distinction between the concrete aspects of vocal regularities, which are described by phonetics, and their abstract representations, described by phonology. Phonetics describes speech as a physical phenomenon and covers the measurable aspects of its acoustic, auditory and articulatory properties. Phonology refers to the abstract representations that allow listeners and speakers to process speech in terms of categorical building blocks that can be grouped, and associated with meaning

(Pierrehumbert, 1990). Thus, one of the important implications that can be drawn from the *Motor Theory of Speech Perception* is that the simulation of the acoustic regularities of language can result in an abstract representation of linguistic units.

### 4.2. Prosody and brain oscillations

The motor theory of speech perception may be considered as a prototype model for the embodiment of linguistic regularities via simulation of sound patterns. However, this model was limited to local regularities involving relatively short segments and therefore it is inadequate for implementing the embodiment of syntactic regularities that span over larger fragments of speech. Moreover, in its original version, the *Motor Theory of Speech Perception* did not explain how the simulation of vocal regularities is implemented in the brain. In what follows we address these issues. We first present the idea that speech analysis involves simultaneous processing of different temporal windows, including long duration windows suitable for processing the prosodic regularities associated with syntax. Then we review evidence from neuro-imaging studies suggesting a mechanism that implements simultaneous processing of different temporal windows. Finally, we focus on the long temporal window required for the processing of phrase level prosody.

### 4.2.1. Temporal windows and neuronal oscillations in speech perception

Although time is often conceived of as a continuous variable, neural systems seem to chunk time in the sense that perceptual information is sampled, integrated and analyzed in discontinuous time windows (Poeppel, 2003). This idea is supported by evidence from psychophysical studies in auditory perception. For example, Saberi and Perrott (1999) showed that speech remained intelligible when the acoustic signal was reversed in chunks of 50 ms, and that although intelligibility dropped with chunks of over 100 ms, it was still above 50%. These findings suggest that information within the relevant time-window is integrated and therefore it can be processed independently of signal direction.

The mechanism suggested for implementing the sampling and integration of information across temporal windows, is the repetitive synchronized behavior of neural populations described as neuronal oscillations. Noninvasive neurophysiological methods such as electroencephalography (EEG) or magnetoencephalography (MEG) reflect the dynamics of neural activity on a millisecond time scale and allow researchers to examine the distribution of neuronal oscillations at different frequencies. Using such methods, it is possible to examine the correlation between neuronal oscillations at different frequency bands and the frequencies of the temporal modulations in the acoustic signal of speech (e.g., Ahissar et al., 2001). In one of the earlier studies that investigated the relationships between neuronal oscillations and speech perception,[4] Ahissar et al. (2001) manipulated speech rate by compressing the acoustic signal while recording speech comprehension and MEG measures of neuronal oscillations. Their findings showed that intelligibility is degraded when activity at the frequency band of 4–8 Hz (theta) ceases to follow the rate of the acoustic modulations in the speech signal due to the compression (see also Luo & Poeppel, 2007; Nourski et al., 2009). These findings demonstrated the functional relationship between neuronal oscillatory rhythms as measured by MEG and the acoustic modulations in speech, and

---

[4] The term speech perception is used differently in different contexts. Here, it does not refer to language comprehension in general (that can be initiated by auditory and visual input), but to the transformation of an auditory signal into mental representations of a type that can make contact with internally stored information (Poeppel & Monahan, 2008).

suggested that the correspondence between them is essential for speech comprehension.

### 4.2.2. Multi-temporal integration windows in speech processing

Ahissar et al.'s (2001) findings further suggest that the frequency of the neuronal oscillations act as temporal integration windows that determine temporal resolution in auditory perception. One of the debates in the literature has focused on the selection of the temporal window most suitable for chunking the acoustic stream for speech analysis. Saberi and Perrott (1999) suggested temporal chunks of 25–50 ms that correspond to acoustic modulations at a frequency of around 30–40 Hz. Nusbaum and Henly (1992) argued that in some situations a smaller window is appropriate, while in others, a larger one is better suited, and therefore suggested adaptive temporal windows.

Critically, the acoustic signal of speech contains information on different time-scales: the rapid changes that distinguish between vowels or between consonants occur on a time scale of 20–40 ms, the changes characteristic of syllables and lexical prosody occur on a time scale of 100–200 ms, and modulations associated with prosodic features such as phrase boundaries or the pitch rise of questions, occur on an even longer time scale. Thus, long duration windows may result in loss of information regarding rapid modulations, whereas short duration windows may result in insensitivity to modulations that span over larger fragments of speech. Based on these observations, it has been suggested that in order to perceive acoustic information at different time scales, simultaneous analysis of different temporal windows is required. Poeppel (2003) proposed a mechanism that may implement multi-scale temporal windows. He hypothesized that different temporal windows of integration are consequences of intrinsic neuronal oscillations at different rates. This hypothesis has been supported by accumulating evidence from neuro-imaging studies.

#### 4.2.2.1. Empirical evidence for multi-temporal integration windows.
Neuro-imaging studies provide both functional and anatomical evidence supporting the idea that stimulus-brain rhythmic alignment across multi-scale temporal windows subserves speech processing over multiple temporal granularities simultaneously. The earlier studies focused mainly on the anatomical aspects of the multi-temporal scales involved in speech analysis. For example Zatorre, Evans, Meyer, and Gjedde (1992) asked participants to listen to the same consonant–vowel–consonant stimulus set, and to perform a judgment task that required either a phonetic discrimination (/t/ vs. /d/) or pitch discrimination (high vs. low tone). Using Positron Emission Tomography (PET) they showed that phonetic judgments were associated with a strong leftward lateralization whereas pitch discrimination was associated with rightward lateralization. These and similar findings suggested hemispheric asymmetries in temporal resolution, with the left hemisphere specializing in rapid acoustic modulations corresponding to phonetic features, and the right hemisphere specializing in slower modulations typical of melodic features (e.g., Zatorre & Belin, 2001; Zatorre, Belin, & Penhune, 2002; Zatorre, Evans, & Meyer, 1994). These asymmetries converge with the findings of Friederici and Alter (2004), showing left hemisphere specialization for the segmental aspects of speech, and right hemisphere specialization for the 'musical' aspects of speech, namely, prosody.

Later studies further revealed that acoustic regularities correspond to rhythmic properties of neuronal oscillations, and showed that this correspondence can be functionally associated with hemispheric asymmetries in the cerebral auditory and motor systems (e.g., Giraud et al., 2007; Luo & Poeppel, 2012). For example, using simultaneous EEG and fMRI, Giraud et al. (2007) showed that fluctuations in brain activity at the high frequency (gamma) range, corresponding to phonemic rate, correlated best with left auditory cortical activity. They also showed that fluctuations within the lower frequency (theta) range, correlated best with cortical activity in the right hemisphere. In addition, they found that fluctuations in both ranges correlate with activity in the pre-motor region controlling the mouth. They interpreted this finding as indicating coupling between temporal properties of speech perception and production (see also Poeppel & Monahan, 2008). Using MEG technology with non-speech auditory stimuli, Luo and Poeppel (2012) provide further support for the anatomical and functional distinction between these two temporal windows in the auditory cortex. Giraud and Poeppel (2012) suggested that this hemispheric asymmetry can be related to cytoarchitectonic differences between the left and right auditory cortex, that induce hemispheric asymmetry in the rhythms of neuronal oscillations in the auditory cortex.

Another line of research that supports the hypothesis of multi-temporal processing windows in speech processing, focused on *inherent* (as opposed to evoked) speech-specialized lateralized brain activity. Morillon et al.'s (2010) data from concurrent EEG and BOLD fMRI recordings show that in the absence of language related processing, oscillatory rhythms at the delta–theta and gamma bands are stronger in left speech-related cortical regions (i.e. auditory, somato-sensory, articulatory motor, and inferior parietal cortices) than in the right homologues regions. While these oscillatory rhythms have been shown to be involved in a range of other cognitive operations, the asymmetry of inherent oscillatory rhythms seems to be confined to the language regions in both hemispheres, suggesting that they play a specialized role in speech processing.

Finally, Telkemeyer et al. (2009, 2011) developmental studies suggest that lateralization of speech perception may be shaped by hemispheric specialization for acoustic properties. For example, Telkemeyer et al. (2009) presented newborns with non-speech auditory stimuli, and showed that fast acoustic modulations corresponding to phonetic rate elicit strong bilateral Auditory Evoked Potential (AEP) responses whereas slow acoustic modulations corresponding to syllabic rate elicited responses lateralized to the right hemisphere. A later study that used Time Frequency Analysis (TFA) further revealed stimulus-induced (as opposed to evoked) changes in oscillatory brain activity, with different patterns of responses to slow compared to fast modulations. Critically, this difference was found in 6 but not in 3 month old infants, suggesting that processing sounds with complex acoustic structure is refined with age, such that perception of slow acoustic modulations develops somewhat later than the perception of fast modulations.

#### 4.2.2.2. Oscillation-based functional model of multi-temporal integration windows.
In view of the accumulating evidence supporting the hypothesis of multi-temporal integration windows, Giraud and Poeppel (2012) suggest that neural oscillations at delta–theta and gamma rhythms serve as the biophysical infrastructure for speech processing. In their *Oscillation-Based Functional model,* they recently proposed that in the process of co-evolution, the motor system involved in speech production and the auditory system involved in speech perception, have tuned to the brain oscillatory rhythms, such that intrinsic oscillations at multi-timescales chunk incoming information into units of temporal granularity suitable for linguistic processing.

According to the *Oscillation-Based Functional model,* intrinsic oscillations at different frequency bands in the auditory cortex interact with neuronal activity generated by the incoming speech signal. This interaction acts as a mechanism of sampling and packaging of the input spike trains that simultaneously generate elementary units at different time scales, corresponding to linguistic units at different levels. In particular, oscillations at the high frequency gamma band (25–35 Hz) correspond to (sub)phonemic units, medium frequency oscillations at the theta band (4–8 Hz)

correspond to syllabic units, and the low-frequency oscillations at the delta (1–3 Hz) band correspond to slower modulations such as phrase level prosody. Moreover, according to the model, intrinsic rhythms of neuronal oscillations and spike timing can be phased locked to generate hierarchic organization of temporal windows. These hierarchically organized windows map the embedding of linguistic units one within another, in the manner that phonemes are embedded within syllables (for the detailed model see Giraud & Poeppel, 2012).

Most of the studies that investigated the hypothesis that speech perception is a process with multi-scale temporal resolution, focused only on two frequency bands - the high (gamma, 25–40 Hz) and medium (theta, 4–8 Hz) frequency bands (but see Cogan & Poeppel, 2011). Nevertheless, theoretical models such as the *Oscillation-Based Functional model* (and see also Ghitza & Greenberg, 2009) assume that there is a third frequency band relevant to speech analysis – the low-frequency delta band (1–3 Hz). This assumption is based mainly on the observation that suprasegmental aspects of the linguistic structure, such as the prosodic modulations associated with phrase boundaries, correspond to longer time scales (Gandour et al., 2003; Rosen, 1992). However, the low frequency oscillations at the delta band haven't been investigated as often as the higher frequency bands. Hence it remained unclear whether the suprasegmental prosodic aspects of speech are processed separately, as reflected in the activity of a distinct frequency band that corresponded to the relevant timescales.

Cogan and Poeppel (2011) addressed this issue in a study that combined MEG recording of oscillatory brain activity and Mutual Information (MI) analysis. MI is an abstract measure (often using bit units) used in information theory to quantify the mutual dependence of two random variables. By adapting MI analysis to MEG methodology, they examined to what degree each frequency band processed independent information from the speech stream. They were particularly interested in the low frequency bands and compared delta (1–3 Hz), theta-low (3–5 Hz), and theta-high (5–7 Hz) frequencies. Their findings indicate that each of these bands processes independent information. Furthermore, using template-base classification analysis, they show that information from each of these frequency bands can be used to classify individual sentences. Classifying results based on combined frequency bands provided the best fit, supporting the conclusion that each frequency band is processing independent information. However it leaves for future research to determine the specific correspondence between the acoustic signal and the neuronal signal. Consequently it is inconclusive with regard to the abstract linguistic units that can be processed by low-frequency oscillations (Cogan & Poeppel, 2011).

We suspect that one of the problems in investigating the correspondence between the rhythms of the neuronal signal and those of prosodic modulations, is the variability in the duration of prosodic modulations. Unlike phonetic or even syllabic units, where the rate of acoustic modulations is relatively stable and well established, prosodic units are highly variable in terms of their temporal duration. Consequently it is hard to determine what the relevant rate of acoustic modulations is, and to examine its correspondence with the rhythms of neuronal oscillations. We believe that future research exploring the systematic associations between prosody and syntax will shed light on the time scale of the relevant prosodic modulations. This will advance the investigation of the correspondence between the prosodic modulations associated with syntax and the relevant rhythms of neuronal oscillations.

### 4.3. Grounding syntax via neuronal oscillations simulating prosodic experience

The findings described above converge to support the view that the *Oscillation-Based model* offers "a link between neurophysiology,

neural computation, acoustics and psycholinguistics" (Giraud & Poeppel, 2012, p. 511). While the *Motor Theory of Speech Perception* focused on relatively short duration segments, Giraud and Poeppel's (2012) *Oscillation-Based model* proposes a multi-time-scale mechanism with extended time windows, namely the delta–theta rhythms, which can capture the slow acoustic modulations underlying prosody. Their model suggests that in the same way that neuronal oscillatory activity at the high frequency gamma band (25–35 Hz) underpins phonetic processing, oscillatory activity at the low frequency delta band (1–3 Hz) underpins prosodic processing.

We posit that neuronal oscillations can serve not only as a mechanism of perceptual analysis but also as the mechanism for simulation of experience. This proposal is in line with the linguistic embodiment view assuming that abstract linguistic representation is implemented via brain activity that simulates experience in the absence of external input. However, unlike classical embodiment models that focus on embodiment via simulation of content and meaning (but see Glenberg & Gallese, 2012) we propose a simulation of linguistic regularities. The temporal nature of neuronal oscillations makes them most suitable for simulating the temporal regularities of the acoustic realization of language – the regularities underlying phonetic, syllabic and prosodic modulations. Building on the *Oscillation-Based model* (Giraud & Poeppel, 2012) we propose that neuronal oscillatory activity at the low-frequency band can serve as the substrate for implementing mental simulation of the prosodic experience.

The appeal of the embodiment approach is that it offers a way to ground abstract linguistic representations. Assuming that neuronal oscillations can simulate the prosodic experience, we can now ask, what abstract representations emerge from this simulation? The answer is complex, as prosody conveys many different types of information, including emotion and social relations. However, we want to focus on linguistic representations. Given the close relationship between prosody and syntax in language processing, we propose that mental simulation of prosodic patterns gives rise to abstract syntactic representations. The evidence suggesting that prosodic features processed in the right hemisphere, and syntactic features processed in the left are coordinated and integrated during speech perception (Sammler et al., 2010), is consistent with our proposal. It seems, then, that the evidence is converging to suggest that prosodic experience simulated by corresponding patterns of neuronal oscillations, offers a mechanism for implementing the embodiment of syntax. Based on the embodiment approach, we assume that such simulation is a necessary prerequisite for syntactic processing and not confined only to challenging conditions. Nevertheless, it is possible that under some conditions (e.g. challenging listening conditions) when prosodic information is incomplete, and/or scarce cognitive resources prevent the prosodic simulation (e.g., articulation suppression during reading, Larsen & Baddeley, 2003) "good enough" comprehension (Ferreira, Bailey, & Ferraro, 2002) can be achieved without the simulation, and consequently without a proper syntactic representation.

## 5. Conclusion

Following the embodiment approach, we argued that in order to gain better understanding of the representation and processing of syntax, we need better understanding of the relationships between syntax, experience and the brain. Traditionally, syntax is conceptualized as the abstract system of linguistic representations underlying the universal human ability to compute nonlinear long range relationships between elements that occur linearly in time such as segmentation, hierarchy and binding. Hence, we first asked

what aspect of the concrete experience of language is most closely related to this abstract system of representations. Given that human language occurs in the acoustic medium, we looked for some aspect of the acoustic experience of language that can be associated with syntax. Reviewing the studies that demonstrate the close relationship between syntax and prosody, suggested that prosody is the acoustic experience most closely associated with syntax. Note that the embodiment of syntax via prosody does not mean that syntax is reduced to prosody. Rather syntax is an emergent linguistic abstraction that can be embodied by different prosodic patterns in different languages. For example, hierarchic relationships between elements are an abstraction that may be coded by intonation in one language and by pause duration in another.

Next we turned to ask what brain activity can be associated with the prosodic experience. Assuming that syntax is more of an organizing than a referential framework, we hypothesized that it is unlikely to be grounded via referential mapping associating content and meaning with activity in related brain regions (e.g. the meaning of '*eating*' embodied by activity in the brain region associated with the mouth). Rather, we proposed that it is more likely to be grounded via non-referential regularities in brain activity. We were looking for brain activity that can simulate the regularities of the acoustic signal that organize information within temporal structures (Kotz & Schwartze, 2010). Review of the studies that investigated the correspondence between regularities in the acoustic signal and neuronal oscillations in the auditory and other speech related brain regions, suggested that neuronal oscillations may provide the means to associate the prosodic experience to brain activity. Building on the *Oscillation-Based model* (Giraud & Poeppel, 2012), we proposed that the neuronal oscillations at the low frequency delta band that were associated with the perception and production of prosody may be elicited for *simulating* prosody, even in the absence of external prosodic input. Such simulation can serve as the substrate for the embodiment of syntax.

Finally, we reviewed evidence indicating that that prosodic features processed in the right hemisphere and syntactic features processed in the left hemisphere are coordinated and integrated during speech perception (e.g., Sammler et al., 2010). These studies demonstrate the functional relationship between brain regions in the right hemisphere that process prosodic features and homologue regions in the left hemisphere involved in syntactic processing. These findings converge with the lateralization pattern revealed in the neuronal oscillation studies. Together they lend support to our proposal that low-frequency oscillations in the right hemisphere, that resonate to perception and production, or to simulation of prosodic regularities, subserve syntactic representations.

### 5.1. Future research

Some of the findings that lend support to our proposal are fairly recent and need to be replicated and established, others are too general to allow the development of a detailed model. We hope that future research inspired by this proposal will address the remaining questions.

For example, although the findings demonstrating the close relationship between syntax and prosody are very robust, they are not very detailed, and they do not reflect cross-linguistic diversity. One problem is that even after half a century of syntactic investigation there is no consensus about what syntax is (Evans & Levinson, 2009), probably because focus has shifted to the extent of cross-linguistic diversity. One of the results of this shift is that it is not clear what aspects of the linguistic experience give rise to syntactic abstractions. In contrast to this ambiguity, prosodic regularities derived from the acoustic signal constitute a physical reality that can be measured. Hence future research that explores

these regularities and reveals their joint variability, together with studies of the cross-linguistic diversity in these regularities, can pave the way to better understanding of the universal vs. language specific syntactic abstractions. Moreover, we believe that future research using advanced signal processing methods for the acoustic analysis of prosody will provide more detailed description of the acoustic modulations underlying particular prosodic features. Such detailed information will improve researchers' ability to explore the correspondence between neuronal oscillations and particular prosodic features, and allow the development of a more detailed model accounting for the embodiment of syntax.

Another issue that needs to be addresses is the somewhat arbitrary division of the frequency continuum into theta, gamma and delta bands (Cogan & Poeppel, 2011). The distinction between the different levels of abstract phonological representations such as phonemic, syllabic and prosodic representations seems to be very clear. Yet it is not clear whether a similar distinction can be made between the corresponding components in the acoustic signal of speech. The notion of oscillatory rhythms that serve as temporal windows of processing offers a mechanism to slice the acoustic stream in a way that should correspond to linguistic representations. However, it is not clear that the frequency bands of neuronal oscillations described above reflect specific "privileged" divisions of the neural signal (Cogan & Poeppel, 2011). Future research should explore the intrinsic rhythms of neuronal oscillations and examine whether they constitute a continuous continuum of rhythms or a final set of privileged bands tuned to the modulation frequency of different linguistic units, as suggested by *Oscillation-Based model* (Giraud & Poeppel, 2012).

Finally, we do not assume that syntactic representations are embodied solely via prosodic simulation. In the same way that classic embodiment models assume that grounding is obtained via multi-modal representation, we assume that syntactic embodiment may be obtained by combined simulation of different regularities, of which prosody plays a major role. It is also possible, for example, that neural circuits involved in controlling the regularities of hierarchy, order, and timing, in motor action also play a role in the embodiment of syntax, as suggested by Glenberg and Gallese (2012). Future research will elucidate the different factors involved in the embodiment of syntax.

Our proposal extends the linguistic embodiment approach and contributes to the dramatic shift from amodal linguistic representation to embodied representation. We extend the embodiment approach in two ways. First, we propose that even syntax, the last sanctuary of abstract linguistic representation, can and should be embodied. Second, we propose a new non-referential concept of grounding, in which abstract syntactic representations emerge from temporal regularities in oscillatory brain activity that simulates the temporal regularities in the acoustic signal. The implication of this shift in emphasis, is that the acoustic signal, and in particular prosody, should have a more central role in the study of brain mechanisms of language processing.

### References

Ahissar, E., Nagarajan, S., Ahissar, M., Protopapas, A., Mahncke, H., & Merzenich, M. (2001). Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. *Proceedings of Natural Academy of Science, 98*, 13367–13372.

Aylett, M., & Turk, A. (2004). The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic

prominence, and duration in spontaneous speech. *Language and Speech, 47*(1), 31–56.

Bader, M. (1998). Prosodic influences on reading syntactically ambiguous sentences. In J. Fodor & F. Ferreira (Eds.), *Reanalysis in sentence processing* (pp. 1–46). Dordrecht, the Netherlands: Kluwer.

Barsalou, L. (2003). Situated simulation in the human conceptual system. *Language and Cognitive Processes, 18*(5–6), 513–562.

Barsalou, L. W. (2008a). Grounded cognition. *Annual Review of Psychology, 59*, 617–645.

Barsalou, L. W. (2008b). Cognitive and neural contributions to understanding the conceptual system. *Current Directions in Psychological Science, 17*(2), 91–95.

Burgess, C., & Lund, K. (1997). Modelling parsing constraints with high-dimensional context space. *Language and Cognitive Processes, 12*(2/3), 177–210.

Byrd, D., Krivokapic, J., & Lee, S. (2006). How far, how long: On the temporal scope of prosodic boundary effects. *The Journal of the Acoustical Society of America, 120*, 1589–1599. http://dx.doi.org/10.1121/1.2217135.

Carlson, K., Frazier, L., & Clifton, C. Jr., (2009). How prosody constrains comprehension: A limited effect of prosodic packaging. *Lingua, 119*, 1066–1082.

Carroll, P. J., & Slowiaczek, M. L. (1987). Modes and modules: Multiple pathways to the language processor. In J. L. Garfield (Ed.), *Modularity in knowledge representation and natural-language understanding* (pp. 221–247). Cambridge, MA, US: MIT Press.

Chomsky, N. (1963). Formal properties of grammars. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. 2, pp. 323–418). New York, London: Wiley.

Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. New York: Harper & Row.

Christophe, A., Mehler, J., & Sebastian-Galles, N. (2001). Perception of prosodic boundary correlates by newborn infants. *Infancy, 2*, 385–394.

Cogan, G. B., & Poeppel, D. (2011). A mutual information analysis of neural coding of speech by low-frequency MEG phase information. *Journal of Neurophysiology, 106*, 554–563. http://dx.doi.org/10.1152/jn.00075.2011.

Cooper, E. W., & Paccia-Cooper, J. (1980). *Syntax and speech*. Cambridge, MA: Harvard University Press.

Cutler, A., Butterfield, S., & Williams, J. N. (1987). The perceptual integrity of syllabic onsets. *Journal of Memory and Language, 26*(4), 406–418.

Cutler, A., Dahan, D., & van Donselaar, W. (1997). Prosody in the comprehension of spoken language: A literature review. *Language and Speech, 40*(2), 141–201.

Dahan, D., Tanenhaus, M. K., & Chambers, C. G. (2002). Accent and reference resolution in spoken-language comprehension. *Journal of Memory and Language, 47*, 292–314.

David, O., Maess, B., Eckstein, B., & Friederici, A. D. (2011). Dynamic causal modeling of subcortical connectivity of language. *The Journal of Neuroscience, 31*(7), 2712–2717.

de Pijper, J. R., & Sanderman, A. A. (1994). On the perceptual strength of prosodic boundaries and its relation to suprasegmental cues. *The Journal of the Acoustical Society of America, 96*, 2037–2047. http://dx.doi.org/10.1121/1.410145.

Eckstein, K., & Friederici, A. D. (2006). It's early: Event-related potential evidence for initial interaction of syntax and prosody in speech comprehension. *Journal of Cognitive Neuroscience, 18*, 1696–1711.

Evans, N., & Levinson, S. C. (2009). The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences, 32*(05), 429–448.

Ferreira, F. (1991). Effects of length and syntactic complexity on initiation times for prepared utterances. *Journal of Memory and Language, 30*(2), 210–233.

Ferreira, F. (1993). Creation of prosody during sentence production. *Psychological Review, 100*, 233–253.

Ferreira, F., Bailey, K. G., & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science, 11*(1), 11–15.

Frazier, L., Carlson, K., & Clifton, C. Jr., (2006). Prosodic phrasing is central to language comprehension. *Trends in Cognitive Science, 10*, 244–249.

Friederici, A. D. (2011). The brain basis of language processing: From structure to function. *Physiological Review, 91*, 1357–1392.

Friederici, A. D., & Alter, K. (2004). Lateralization of auditory language functions: A dynamic dual pathway model. *Brain & Language, 89*, 267–276.

Friederici, A. D., von Cramon, D. Y., & Kotz, S. A. (2007). Role of the corpus callosum in speech comprehension: Interfacing syntax and prosody. *Neuron, 53*(1), 135–145.

Fodor, J. D. (2002). Psycholinguistics cannot escape prosody. In B. Bel & I. Marlien (Eds.), *Proceedings of the 1st international conference on speech prosody* (pp. 83–88). University of Tokio, Japan: Keikichi Hirose Laboratory.

Galantucci, B., Fowler, C. A., & Turvey, M. T. (2006). The Motor Theory of speech perception reviewed. *Psychonomic Bulletin & Review, 13*(3), 361–377.

Gandour, J., Dzemidzic, M., Wong, D., Lowe, M., Tong, Y., Hsieh, L., et al. (2003). Temporal integration of speech prosody is shaped by language experience: An fMRI study. *Brain and Language, 84*, 318–336.

Ghitza, O. (2011). Linking speech perception and neurophysiology: Speech decoding guided by cascaded oscillators locked to the input rhythm. *Frontiers in Psychology, 2*, 130.

Ghitza, O., & Greenberg, S. (2009). On the possible role of brain rhythms in speech perception: Intelligibility of time-compressed speech with periodic and aperiodic insertions of silence. *Phonetica, 66*, 113–126. http://dx.doi.org/10.1159/000208934.

Gibbs, R. W. Jr., (2003). Embodied experience and linguistic meaning. *Brain and Language, 84*, 1–15.

Giraud, A. L., Kleinschmidt, A., Poeppel, D., Lund, T. E., Frackowiak, R. S. J., & Laufs, H. (2007). Endogenous cortical rhythms determine cerebral specialization for speech perception and production. *Neuron, 56*, 1127–1134.

Giraud, A. L., & Poeppel, D. (2012). Cortical oscillations and speech processing: Emerging computational principles and operations. *Nature Neuroscience, 15*(4), 511–517.

Glenberg, A. M. (2010). Embodiment as a unifying perspective for psychology. *Wiley Interdisciplinary Reviews: Cognitive Science, 1*(4), 586–596.

Glenberg, A. M., & Gallese, V. (2012). Action-based language: A theory of language acquisition, comprehension, and production. *Cortex, 48*, 905–922.

Goldberg, A. E. (2003). Constructions: A new theoretical approach to language. *Trends in Cognitive Sciences, 7*(5), 219–224.

Grodzinsky, Y. (2006). A blueprint for a brain map of syntax. In Y. Grodzinsky & K. Amunts (Eds.), *Broca's region*. New York: Oxford University Press.

Grodzinsky, Y. (2010). The picture of the linguistic brain: how sharp can it be? Reply to Fedorenko & Kanwisher. *Language and Linguistics Compass, 4*(8), 605–622.

Grosjean, F. H., Grosjean, L., & Lane, H. (1979). The patterns of silence: Performance structures in sentence production. *Cognitive Psychology, 11*(1), 58–81.

Harris, C. M. (1953). A study of the building blocks in speech. *Journal of Acoustic Society of America, 25*, 962–969.

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience, 8*(5), 393–402.

Hirschberg, J., & Ward, G. (1991). Accent and bound anaphora. *Cognitive Linguistics, 2*, 101–121.

Hwang, H., & Steinhauer, K. (2011). Phrase length matters: The interplay between implicit prosody and syntax in Korean "Garden Path" sentences. *Journal of Cognitive Neuroscience, 23*(11), 3555–3575.

Johnson, E., & Jusczyk, P. W. (2001). Word segmentation by 8 month olds: When speech cues count more than statistics. *Journal of Memory & Language, 44*, 548–567.

Jusczyk, P. W., Hirsh-Pasek, K., Kemler-Nelson, D., Kennedy, L., Woodward, A., & Piwoz, J. (1992). Perception of acoustic correlates of major phrasal units by young infants. *Cognitive Psychology, 24*, 252–293.

Kelly, M. H. (1992). Using sound to solve syntactic problems: The role of phonology in grammatical category assignments. *Psychological Review, 99*(2), 349–364.

Kintsch, W., & Van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review, 85*, 363–394.

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

Klouda, G. V., Robin, D. A., Graff-Radford, N. R., & Copper, W. E. (1988). The role of callosal connections in speech prosody. *Brain & Language, 35*, 154–171.

Kotz, S. A., & Schwartze, M. (2010). Cortical speech processing unplugged: A timely subcortico-cortical framework. *Trends in Cognitive Science, 14*(9), 392–399.

Kremers, J. (2012). The syntax of simultaneity. *Lingua, 122*(9), 979–1003.

Ladd, D. R. (2008). *Intonational phonology* (2nd ed.). Cambridge, UK: Cambridge University Press.

Lakoff, G., & Johnson, M. (2008). *Metaphors we live by*. University of Chicago Press.

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.

Larsen, J. D., & Baddeley, A. (2003). Disruption of verbal STM by irrelevant speech, articulatory suppression, and manual tapping: Do they have a common source? *Quarterly Journal of Experimental Psychology Section A, 56*(8), 1249–1268.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of speech code. *Psychological Review, 74*, 431–461.

Liberman, A. M., & Whalen, D. H. (2000). On the relation of speech to language. *Trends in Cognitive Science, 4*, 187–196.

Lieberman, P. (2007). The evolution of human speech: Its anatomical and neural bases. *Current Anthropology, 48*(1), 39–66.

Luo, H., & Poeppel, D. (2007). Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron, 54*(6), 1001–1010.

Luo, H., & Poeppel, D. (2012). Cortical oscillations in auditory perception and speech: Evidence for two temporal windows in human auditory cortex. *Frontiers in Psychology, 3*, 170–181. http://dx.doi.org/10.3389/fpsyg.2012.00170.

Mahon, B. Z., & Caramazza, A. (2009). Concepts and categories: A cognitive neuropsychological perspective. *Annual Review of Psychology, 60*, 27–51.

Mandel, D. R., Jusczyk, P., & Nelson, D. (1994). Does sentential prosody help infants organize and remember speech information? *Cognition, 53*, 155–180.

Mattingly, I. G., & Liberman, A. M. (1988). Specialized perceiving systems for speech and other biologically significant sounds. In G. M. Edelman, W. E. Gall, & W. M. Cowan (Eds.), *Auditory function: Neurobiological bases of hearing* (pp. 775–793). New-York: John Wiley & Sons.

Mehler, J., Jusczyk, P. W., Lambertz, G., Halsted, N., Bertoncini, J., & Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition, 29*, 143–178.

Morillon, B., Lehongre, K., Frackowiak, R. S. J., Ducorps, A., Kleinschmidt, A., Poeppel, D., et al. (2010). Neurophysiological origin of human brain asymmetry for speech and language. *Proceedings of Natural Academy of Science, 107*(43), 18688–18693.

Nelson, K. (2003). Narrative and the emergence of a consciousness of self. In G. Firenam, Jr., T. E. McVay, Jr., & O. J. Flanagan (Eds.), *Narrative and consciousness: Literature, psychology, and the brain* (pp. 17–50). Oxford, UK: Oxford University Press.

Nespor, M., & Vogel, I. (1986). *Prosodic phonology* (1st ed.). Berlin, Germany: Mouton de Gruyter [Dordrecht, Foris].

Niedenthal, P. M. (2007). Embodying emotion. *Science, 316*(5827), 1002–1005.

Nourski, K. V., Reale, R. A., Oya, H., Kawasaki, H., Kovach, C. K., Chen, H., et al. (2009). Temporal envelope of time-compressed speech represented in the human auditory cortex. *Journal of Neuroscience, 29*, 15564–15574.

Nusbaum, H. C., & Henly, A. S. (1992). Constraint satisfaction, attention, and speech perception: Implications for theories of word perception. In M. E. H. Schouten (Ed.), *The auditory processing of speech: From sounds to words*. Berlin, Germany: Mouton.

Pierrehumbert, J. (1990). Phonological and phonetic representation. *Journal of Phonetics, 18*, 375–394.

Poeppel, D. (2003). The analysis of speech in different temporal integration windows: Cerebral lateralization as 'asymmetric sampling in time'. *Speech Communication, 41*, 245–255.

Poeppel, D., & Monahan, P. J. (2008). Speech perception: Cognitive foundations and cortical implementation. *Current Directions in Psychological Science, 17*(2), 80–85.

Pollmann, S., Maertens, M., von Cramon, D. Y., Lepsien, J., & Hughdahl, K. (2002). Dichotic listening in patients with posterior CC and nonsplenial callosal lesions. *Neuropsychologia, 16*, 56–64.

Price, P., Ostendorf, M., Shattuck-Hufnagle, S., & Fong, C. (1991). The use of prosody in syntactic disambiguation. *Journal of the Acoustical Society of America, 90*, 2956–2970.

Prinz, J. (2002). *Furnishing the mind: Concepts and their perceptual basis*. Cambridge, MA: MIT.

Pulvermüller, F. (2005). Brain mechanisms linking language and action. *Nature Reviews Neuroscience, 6*(7), 576–582.

Pulvermüller, F. (2010). Brain embodiment of syntax and grammar: Discrete combinatorial mechanisms spelt out in neuronal circuits. *Brain & Language, 112*, 167–179.

Pulvermüller, F., Huss, M., Kheri, F., Moscoso del Prado Martin, F., Hauk, O., & Shtyrov, Y. (2006). Motor cortex maps articulatory features of speech sounds. *Proceedings of the National Academy of Sciences, 103*, 7865–7870.

Pulvermüller, F., & Fadiga, L. (2010). Active perception: Sensorimotor circuits as a cortical basis for language. *Nature Reviews Neuroscience, 11*, 351–360.

Pynte, J. (1996). Prosodic breaks and attachment decisions in sentence parsing. *Language and Cognitive Processes, 11*(1–2), 165–192.

Rosen, S. (1992). Temporal information is speech: Acoustic, auditory, and linguistic aspects. *Philosophical Transactions of the Royal Society of London, 336*, 367–373.

Saberi, K., & Perrott, D. R. (1999). Cognitive restoration of reversed speech. *Nature, 398*, 760.

Sammler, D., Kotz, S. A., Eckstein, K., Ott, D. V. M., & Friederici, A. D. (2010). Prosody meets syntax: The role of the corpus callosum. *Brain, 133*, 2643–2655.

Sandler, W., Meir, I., Dachkovsky, S., Padden, C., & Aronoff, M. (2011). The emergence of complexity in prosody and syntax. *Lingua, 121*, 2014–2033.

Selkirk, E. (1984). *Phonology and syntax: The relation between sound and structure*. Cambridge, MA: The MIT Press.

Smith, L. B. (2005). Cognition as a dynamic system: Principles from embodiment. *Developmental Review, 25*(3), 278–298.

Snedeker, J., & Casserly, E. (2010). Is it all relative? Effects of prosodic boundaries on the comprehension and production of attachment ambiguities. *Language and Cognitive Processes, 25*(7–9), 1234–1264.

Speer, S. R., & Blodgett, A. (2006). Prosody. In M. Traxler & M. A. Garnsbacher (Eds.), *Handbook of psycholinguistics* (2nd ed., pp. 505–537). Amsterdam, The Netherlands: Elsevier.

Speer, S. R., Crowder, R. G., & Thomas, L. M. (1993). Prosodic structure and sentence recognition. *Journal of Memory and Language, 32*(3), 336–358.

Speer, S. R., & Ito, K. (2009). Prosody in first language acquisition – Acquiring intonation as a tool to organize information in conversation. *Language and Linguistics Compass, 3*(1), 90–110.

Speer, S., Warren, P., & Schafer, A. J. (2011). Situationally independent prosody. *Laboratory Phonology, 2*, 35–98.

Steedman, M. (2000). Information structure and the syntax-phonology interface. *Linguistic Inquiry, 31*(4), 649–689.

Steinhauer, K., & Friederici, A. (2001). Prosodic boundaries, comma rules, and brain responses: The closure positive shift in ERPs as a universal marker for prosodic phrasing in listeners and readers. *Journal of Psycholinguistic Research, 30*(3), 267–295.

Telkemeyer, S., Rossi, S., Koch, S. P., Nierhaus, T., Steinbrink, J., Poeppel, D., et al. (2009). Sensitivity of newborn auditory cortex to the temporal structure of sounds. *The Journal of Neuroscience, 29*(47), 14726–14733.

Telkemeyer, S., Rossi, S., Nierhaus, T., Steinbrink, J., Obrig, H., & Wartenburger, I. (2011). Acoustic processing of temporally modulated sounds in infants: Evidence from a combined near-infrared spectroscopy and EEG study. *Frontiers in Psychology, 2*(62), 1–14.

Tettamanti, M., & Moro, A. (2012). Can syntax appear in a mirror (system)? *Cortex, 48*, 923–935.

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.

Truckenbrodt, H. (2002). Upstep and embedded register levels. *Phonology, 19*(1), 77–120. http://dx.doi.org/10.1017/S095267570200427X.

Wagner, M., & Watson, D. G. (2010). Experimental and theoretical advances in prosody: A review. *Language and Cognitive Processes, 25*, 905–945.

Watson, D. G., & Gibson, E. A. (2004). The relationship between intonational phrasing and syntactic structure in language production. *Language and Cognitive Processes, 19*(6), 713–755.

Willems, R. M., & Hagoort, P. (2007). Neural evidence for the interplay between language, gesture, and action: A review. *Brain and Language, 101*, 278–289.

Wilson, S. M., Saygin, A. P., Sereno, M. I., & Iacoboni, M. (2004). Listening to speech activates motor areas involved in speech production. *Nature Neuroscience, 7*, 701–702.

Yao, B., Belin, P., & Scheepers, C. (2011). Silent reading of direct versus indirect speech activates voice-selective areas in the auditory cortex. *Journal of Cognitive Neuroscience, 23*(10), 3146–3152.

Zatorre, R. J., Evans, A. C., & Meyer, E. (1994). Neural mechanisms underlying melodic perception and memory for pitch. *Journal of Neuroscience, 14*, 1908–1919.

Zatorre, R., Evans, A., Meyer, E., & Gjedde, A. (1992). Lateralization of phonetic and pitch discrimination in speech processing. *Science, 256*, 846–849.

Zatorre, R. J., Belin, P., & Penhune, V. B. (2002). Structure and function of auditory cortex: Music and speech. *Trends in Cognitive Science, 6*, 37–46.

Zatorre, R. J., & Belin, P. (2001). Spectral and temporal processing in human auditory cortex. *Cereral Cortex, 11*(10), 946–953.

Zwaan, R. A. (2008). Time in language, situation models, and mental simulations. *Language Learning, 58*, 13–26.