brain sciences

MDPI

Review

Pushing the Envelope: Developments in Neural Entrainment to Speech and the Biological Underpinnings of Prosody Perception

Brett R. Myers [1,2,*], **Miriam D. Lense** [1,3,4,5] **and Reyna L. Gordon** [1,4,5,6,*]

[1] Department of Otolaryngology, Vanderbilt University Medical Center, 1215 21st Ave S, Nashville, TN 37232, USA
[2] Department of Psychology and Human Development, Vanderbilt University, 230 Appleton Place, Nashville, TN 37203, USA
[3] Vanderbilt Kennedy Center, 110 Magnolia Circle, Nashville, TN 37203, USA; miriam.lense@vumc.org
[4] Vanderbilt Brain Institute, Vanderbilt University, 2215 Garland Ave, Nashville, TN 37232, USA
[5] The Curb Center for Art, Enterprise, and Public Policy, Vanderbilt University, 1801 Edgehill Avenue, Nashville, TN 37212, USA
[6] Department of Psychology, Vanderbilt University, 2301 Vanderbilt Place, Nashville, TN 37240, USA
* Correspondence: brett.myers@vanderbilt.edu (B.R.M.); reyna.gordon@vanderbilt.edu (R.L.G.)

Received: 31 December 2018; Accepted: 15 March 2019; Published: 22 March 2019

Abstract: Prosodic cues in speech are indispensable for comprehending a speaker's message, recognizing emphasis and emotion, parsing segmental units, and disambiguating syntactic structures. While it is commonly accepted that prosody provides a fundamental service to higher-level features of speech, the neural underpinnings of prosody processing are not clearly defined in the cognitive neuroscience literature. Many recent electrophysiological studies have examined speech comprehension by measuring neural entrainment to the speech amplitude envelope, using a variety of methods including phase-locking algorithms and stimulus reconstruction. Here we review recent evidence for neural tracking of the speech envelope and demonstrate the importance of prosodic contributions to the neural tracking of speech. Prosodic cues may offer a foundation for supporting neural synchronization to the speech envelope, which scaffolds linguistic processing. We argue that prosody has an inherent role in speech perception, and future research should fill the gap in our knowledge of how prosody contributes to speech envelope entrainment.

Keywords: prosody; speech envelope; neural entrainment; rhythm; EEG

> "In a house constructed of speech, the bricks are phonemes, and the mortar is prosody. Without the latter, we'd simply live under a pile of rocks". —B.R.M.

1. Prosody Perception

Prosody is the stress, intonation, and rhythm of speech, which provides suprasegmental linguistic features across phonemes, syllables, and phrases [1–3]. Prosodic cues contribute affect and intent to an utterance [4] as well as emphasis, sarcasm, and more nuanced emotional states [5,6]. Certain prosodic cues are universal and can be interpreted cross-culturally even in an unfamiliar language [7,8]. Prosody also provides valuable markers for parsing a continuous speech stream into meaningful segments such as intonational phrase boundaries [9], dynamic pitch changes [10], and metrical information [11]. Parsing speech units based on prosodic perception is an imperative early stage in language acquisition, and it is considered a precursor to vocabulary and grammar development [12–14]. In addition, prosody can convey semantic information for context in a message [15,16]. Deficits in prosody perception have a negative downstream impact on linguistic abilities, literacy, and social interactions, e.g., [17–20].

Prosodic fluctuations are responsible for communicating a wealth of information, primarily through acoustic correlates such as duration, amplitude, and fundamental frequency. As any of these parameters changes, it influences the expression of stress, intonation, and rhythm of the spoken message [21,22]. One illustration of the dynamic and multidimensional nature of prosody is "motherese" or infant-directed speech, which is characterized by exaggerations in duration and fundamental frequency [23]. The exaggerated speech signal creates louder, longer, and higher pitch stressed syllables [24], which facilitates segmenting the speech into syllable components and disentangling word boundaries [25,26]. The modified prosodic qualities of infant-directed speech make the signal acoustically salient and engaging for infants [23,27], which yield later linguistic benefits such as boosts in vocabulary acquisition [28] and accessing syntactic structures [29]. This is one example of how prosody plays an important role in speech communication.

The importance of prosody to speech perception is widely acknowledged, yet it has been underrated in many studies examining neural entrainment to the speech envelope. The purpose for the current review is to demonstrate that prosodic processing is engrained in investigations of neural entrainment to speech and to encourage researchers to explicitly consider the effects of prosody in future investigations. We will review the speech envelope and its relation to the prosodic features of duration, amplitude, and fundamental frequency, and we will discuss electrophysiological methods for measuring speech envelope entrainment in neural oscillations. We will then highlight some previous research using these methods in typical and atypical populations with an emphasis on how the findings may be connected to prosody. Finally, we propose directions for future research in this field. It is our hope to draw attention to the role of prosody processing in neural entrainment to speech and to encourage researchers to examine the neural underpinnings of prosodic processing.

2. Amplitude Modulation

Prosody is determined by a series of acoustic correlates—duration, amplitude, and fundamental frequency—which can be represented in a number of ways, including the amplitude modulation (AM) envelope (also known as the temporal envelope) [30,31]. It is important to mention that a temporal waveform is composed of a "fine structure" and an "envelope". Fine structure consists of fast-moving spectral content (e.g., frequency characteristics of phonemes), while the envelope captures the broad contour of pressure variations in the signal (e.g., amplitude over time) [32]. In other words, the envelope is superimposed over the more rapidly oscillating fine structure. Both envelope and spectral components are important for speech comprehension—i.e., to "recognize speech" rather than "wreck a nice beach" (Figure 1) (see [33] but also [34]).

It has been suggested that the extraction of AM information is a fundamental procedure within the neural architecture of the auditory system [35]. The auditory cortex is particularly adept at rapidly processing spectro-temporal changes in the temporal fine structure [36], and it is possible that this processing is aided by the amplitude envelope first laying the foundation for more narrow linguistic structure [37]. For example, fine structure cues play an important role in speech processing, yet normal-hearing listeners are able to detect these cues from envelope information alone [38]. Even when spectral qualities are severely degraded, speech processing can be achieved with primarily envelope information [39], as the envelope provides helpful cues for parsing meaningful segments in speech [40,41]. Additionally, the temporal characteristics of an auditory object allow us to focus attention on the source and segregate it from competing sources [42], which makes detection of envelope cues essential in speech communication. Because the amplitude envelope captures suprasegmental features across the speech signal, it lends itself to being an excellent proxy for prosodic information, and we argue that studies that use the speech envelope are inherently targeting a response to prosody.

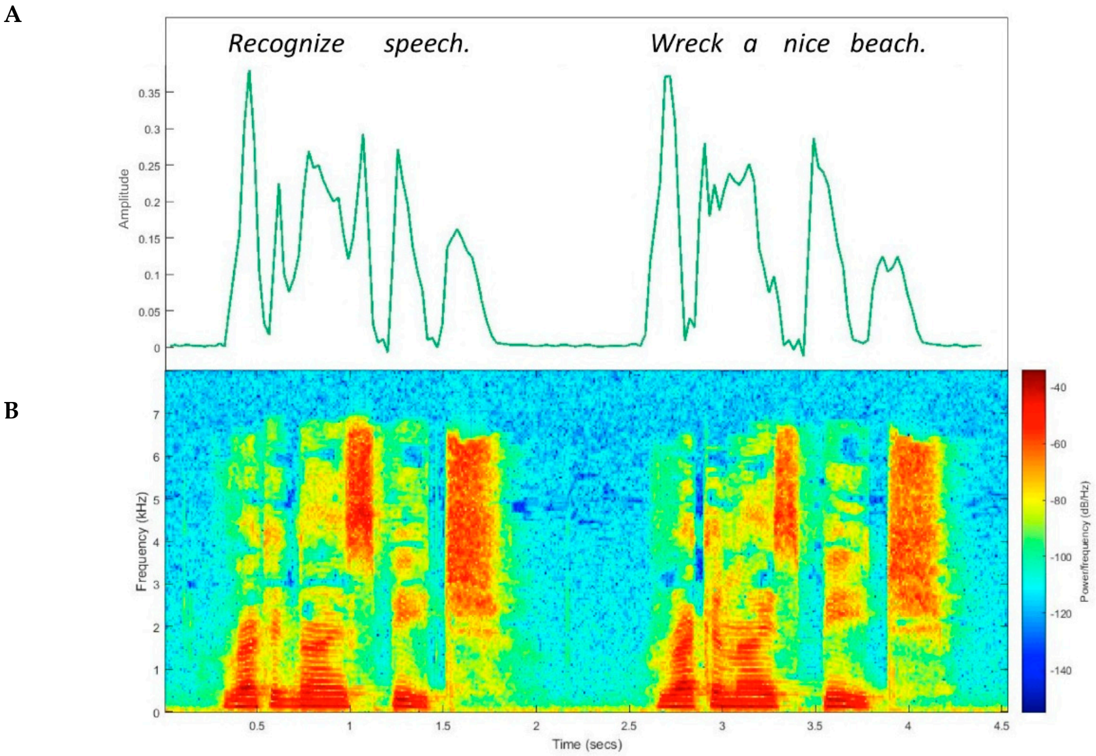

Figure 1. Two representations of an acoustic speech signal: Amplitude envelope (**A**) and spectrogram (**B**). Subtle differences between the phrases "recognize speech" and "wreck a nice beach" can be detected in both representations.

The speech amplitude envelope provides a linear representation of AM fluctuations over time. Acoustic stimuli are constructed of multiple temporal dimensions [31], and modulation energy varies based upon the selected band of carrier frequencies in the signal [35]. Speech can be portrayed through a hierarchical series of AM frequency scales [43]; that is, stress placement occurs at a rate of ~2 Hz [44], syllable rate occurs around 3–5 Hz [24], and phonemic structure has a faster rate of 8–50 Hz [31] (see Figure 2). Liss et al. [45] found that energy in the frequency bands below 4 Hz was intercorrelated, and energy above 4 Hz (up to 10 Hz) was separately intercorrelated. The frequency range between 4 and 16 Hz primarily affects speech intelligibility [46], while frequencies below 4 Hz strictly reflect prosodic variations, such as stress and syllable rate [47]. These multiple timescales of modulation energy within the speech envelope have been shown to elicit corresponding modulations in cortical activity during speech processing [48]. This correspondence appears to play a role in potentially challenging listening situations, such as: speech in noise [49], multiple speakers [50], complex auditory scenes [51], conflicting visual information [52], and divided attention [53,54]. Each of these situations (discussed in more detail later) requires the listener to exploit the natural timing of speech using prosodic cues, which are provided in the amplitude envelope [55].

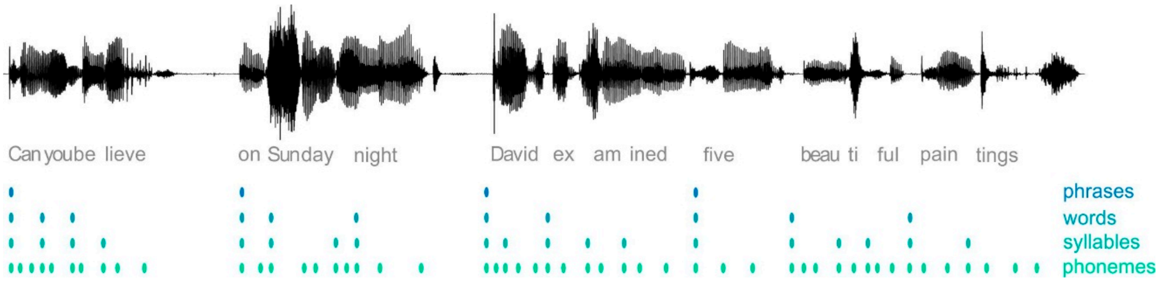

Figure 2. Acoustic waveform with its segmentation into phrases, words, syllables, and phonemes. Figure reproduced from [48].

3. Neural Entrainment to the Speech Envelope

Neural entrainment to the speech envelope has been a notoriously complex topic of study for several decades. In this section we will provide a broad overview of some investigative strides in this area. It is well known that neural oscillatory activity occurs in a constant stream of peaks and troughs while at rest and during cognitive processes. This stream becomes an adaptive spike train in response to environmental stimuli, such as the acoustic signal of speech. Numerous studies have shown that neural oscillatory activity in specific frequency bands is related to specific linguistic functions; for example, lower-level linguistic processing, such as detection of stress and syllable segmentation, occurs in lower frequencies (<4 Hz) [47], and semantic/syntactic processing may occur in higher frequencies (13–50 Hz) [56].

Traditional EEG approaches to prosody perception include analyzing event-related potential (ERP) activity at key events in the speech signal [57], such as stressed syllables [58], metric structure violations [59], pitch violations [60], and duration violations [11]. While these techniques are important for determining brain responses to prosodic features, they do not provide a comprehensive measure of how the brain tracks and encodes the multidimensional aspects of prosody over time. Because prosody refers to suprasegmental features (duration, amplitude, fundamental frequency), which vary throughout an utterance, it is useful to analyze prosody across the temporal domain rather than at one point in time. For this we turn to the speech amplitude envelope as a representation of suprasegmental information.

Recent developments in the literature have explored ways to measure continuous neural entrainment, which is a phenomenon where neuronal activity synchronizes to the periodic qualities of the incoming stimuli [61]. The oscillations of the auditory cortex reset their phase to the rhythm of the speech signal, which is an essential process for speech comprehension [33]. This is known as phase-locking, which can be measured with a cross-correlation procedure between the speech stimulus and the resultant M/EEG signal [62]. Cross-correlation uncovers similarities between two time series using a range of lag windows [63]. This is an efficient method for observing the response to continuous speech without requiring a large number of stimulus repetitions, since this analysis inherently increases the signal-to-noise ratio [64].

Speech processing occurs through a large network of cortical sources [65], and phase-locking can be measured to locate functionally independent sources [66]. These sources may occur bilaterally depending on the timescale [67], such that the left hemisphere favors rapid temporal features of speech, and the right hemisphere tracks slower features. The right hemisphere generally shows stronger tracking of the speech envelope [68,69]; however, envelope tracking has also been shown to be a bilateral process [62,70].

When measuring how the speech envelope is represented in neural data, one issue with a simple cross-correlation between envelope and neural response is that temporal smearing (from averaging across time points) will create noise in the correlation function [71]. A solution to this is to use a modeling approach, known as a temporal response function (TRF) [68], to describe the linear mapping between stimulus and response. This approach stems from a system identification technique [72] that models the human brain as a linear time-invariant system. Of course, the brain does not operate on a linear or time-invariant schedule, but these assumptions are commonly accepted in neurophysiology research for characterizing the system by its impulse response [73,74].

The modeling approach can operate in either the forward or backward direction. Forward modeling describes the mapping of a speech stimulus to a neural network [68,75,76] using a TRF that represents the linear transformation that generated the observed neural signal [77] (Figure 3). When using the envelope representation of speech, the forward model treats the stimulus as a univariate input affecting each recording channel separately. However, since the speech signal is transformed in the auditory pathway into multiple frequency bands [78], the forward modeling procedure may benefit from a multivariate temporal response function (mTRF) [71], which uses the spectrogram representation to evaluate speech encoding. Even in the multivariate domain, forward modeling still maps the stimulus to each response channel independently [79].

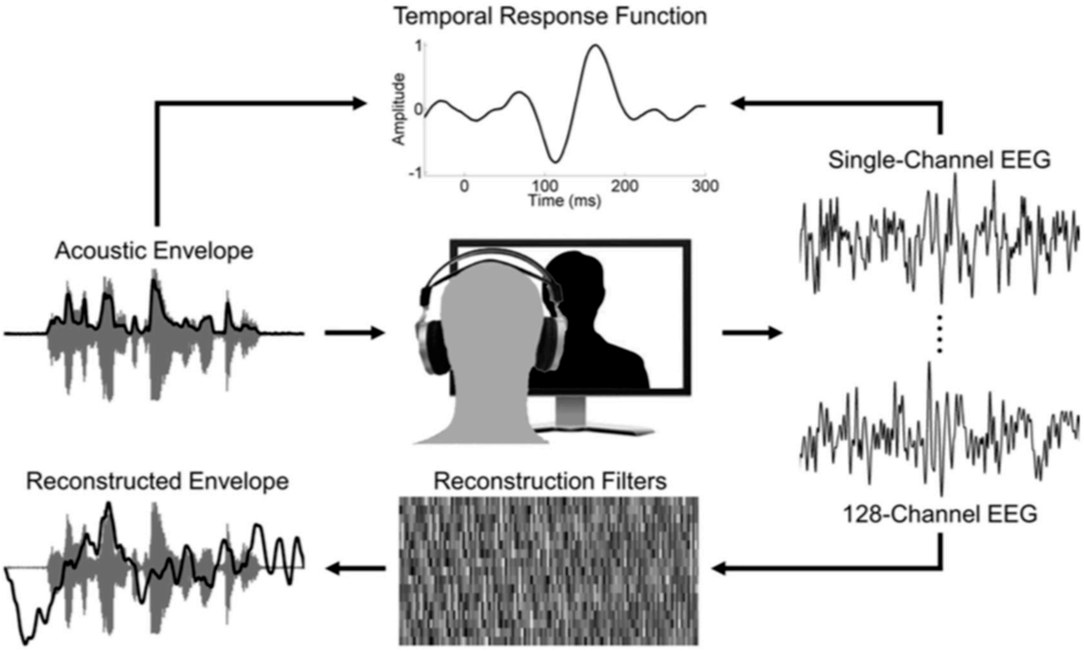

Figure 3. The temporal response function (TRF)—calculated with a linear least squares approach—represents the mapping from acoustic envelope onto each channel of EEG data (forward modeling). A multivariate reconstruction filter can be applied to data from all channels to estimate the acoustic envelope (backward modeling). Reconstruction accuracy can be measured by Pearson correlation between original and reconstructed envelopes. Figure reproduced from [75].

Backward modeling is a mathematical representation of the linear mapping from the multivariate neural response back to the stimulus [71]. This modeling approach yields a decoder that attempts to reconstruct a univariate stimulus feature, such as the speech envelope. As described in [71], this decoder function is derived by minimizing the mean squared error between the stimulus and reconstruction. In the backward direction, recording channels are weighted based on the information that they provide for the reconstruction [77], which removes inter-channel redundancies—an advantage over forward modeling. By modeling in the backward direction, researchers are able to compare stimulus reconstructions to the original stimulus, for instance with a correlation coefficient as a marker of reconstruction accuracy [80]. This provides a reliable index for the degree to which the envelope is encoded in the neural network. While other methods—such as cross-correlations and inter-trial phase coherence—are adequate for measuring phase-locking in speech comprehension, the modeling approach has been gaining attention as an attractive analysis method in recent years. Regardless of the method used, measuring neural entrainment to the speech envelope is an excellent way to target prosodic processing, yet this has been underutilized in the literature.

4. Selected Findings in Envelope Entrainment

Many questions about speech processing can be investigated by looking at neural entrainment to the speech envelope, though we must be careful about how we interpret the results (see [34]; Table 1 provides a summary of selected studies). Peelle et al. [32] compared intelligible speech with unintelligible noise-vocoded speech, and they found that cortical oscillations in the theta (4–7 Hz) band are more closely phase-locked to intelligible speech. This may suggest that linguistic information and contextual associations enhance phase-locking to the envelope. However, others have measured envelope tracking in the auditory cortex even when the signal is devoid of communicative value. For example, Nourski et al. [81] found envelope entrainment even when speech rate was compressed to an unintelligible degree; Howard and Poeppel [82] found envelope entrainment to time-reversed speech stimuli; Mai et al. [56] found envelope entrainment to pseudo-word utterances. We acknowledge that envelope entrainment is often enhanced by intelligibility [49], but given the conflicting results described here, it is difficult to say whether intelligibility predicts entrainment or vice versa. What we can take away from these findings is that acoustic features of the stimulus—such as suprasegmental cues—seem to contribute to the neural entrainment effect and that the effect of neural entrainment on speech intelligibility warrants further investigation.

Attention has also been shown to influence envelope entrainment, and selective attention in a multi-speaker environment can be observed by the degree to which neural oscillations entrain to a given speech envelope [54,61,83]. The classic cocktail party situation has been studied for decades [84] and continues to be of interest today, e.g., [85]. In a natural auditory environment, many sounds are merged together and presented to the ear simultaneously, and the listener is tasked with segregating the sounds and attending to a particular source while ignoring the others [86]. By analyzing speech envelope representations, we can determine how the neural circuit parses and segregates these auditory objects. Ding and Simon [54] demonstrated that when a listener hears two speakers simultaneously, the neural decoding process is able to reconstruct the stimulus envelopes of both speech streams. The stimulus reconstruction is more strongly correlated to the envelope of the attended speaker (also [61,62,83]). Similar results have also been shown with invasive electrodes in electrocorticography (ECoG) research [53,87]. Despite the methodology used, these studies have suggested that neural encoding of an auditory scene involves selective phase-locking to specific auditory objects that are presented concurrently in a single auditory mixture. As mentioned previously, prosodic features of a speech stream help a listener to parse speech and attend to it, so prosody likely plays an important role in multi-speaker envelope entrainment, yet manipulations of the prosodic features of speech are rarely included as a variable in multi-speaker entrainment studies.

Speech envelopes are also of interest in studies examining audiovisual presentation of speech. Visual speech provides critical information regarding the timing and content of the acoustic signal [88]. It has long been acknowledged that listeners perceive speech better when they can both see and hear the individual speaking [89]. Articulatory and facial movements provide visual temporal cues that complement meaningful markers in the auditory stream. Visual rhythmic movements help parse syllabic boundaries [90], a wider mouth opening indicates louder amplitude [91], and seeing a conversational partner assists in segregating a speech stream from overlapping speakers [92]. Visual cues and gestures are tightly linked to speech prosody [93–95], and this alignment emphasizes suprasegmental features of the speech signal.

When auditory and visual information are incongruous, speech perception may be hindered and even lead the listener to falsely perceive a sound that was not presented in either modality (à la "The McGurk Effect") [96]. Congruent audiovisual speech enhances envelope tracking compared to incongruent information and also shows greater envelope encoding than auditory only speech, visual only speech, or the combination of the two unisensory modalities [97]. Audiovisual speech also has marked benefits for neural tracking when presented in noisy conditions [98] (also [99,100]). This is indicative of multisensory enhancement during speech envelope encoding. At the same time, there appears to be a similar mechanism for visual entrainment in which cortical oscillations entrain to salient lip movements even when they are incongruous to the acoustic stream [101]. These studies of envelope responses to speech incongruence support an emerging model of correlated auditory and visual signals dynamically interacting in a discrete process of multisensory integration [88,102]. Prosody is a major factor in this integration, as it aligns a stable framework of temporal and acoustic–phonetic cues to be used in speech processing; however, the contribution of prosodic dimensions of the speech stimuli to neural entrainment in multisensory processing in these studies has not been explicitly considered.

Prosody shares a number of features with music, so an area for potential exploration is the connection between neural entrainment to speech and to music. Envelope entrainment is influenced by speech rhythm [103]. Because rhythm and temporal cues provide a common link between music and speech perception (e.g., [104,105]), several studies demonstrate associations between musical rhythm aptitude, speech perception, and literacy skills in children [106,107]. Some have hypothesized that entrainment to music leads to increased timing precision in the auditory system, which leads to increased perception of the timing of speech sounds [108,109]. Doelling and Poeppel [110] found that the accuracy of cortical entrainment to musical stimuli is contingent upon musical expertise, suggesting individual differences in cortical oscillations related to experience. However, musical expertise does not necessarily predict stronger entrainment to the speech envelope [111]. Additional work on individual differences between speech and music may help to target the neural mechanisms behind prosodic processing.

Table 1. List of papers investigating speech envelope tracking using various analysis approaches, data collection procedures, and topics of interest. Analysis abbreviations: CC—cross-correlation; PC—phase coherence; TRF—temporal response function; SR—stimulus reconstruction.

Author/Year	Data	Analysis	Relevant Amplitude Envelope Findings
Speech Intelligibility			
Ahissar et al., 2001 [63]	MEG	CC	Phase-locking predicts speech comprehension
Luo and Poeppel, 2007 [112]	MEG	PC	Phase-locking to speech is robust at 4–8 Hz
Abrams et al., 2008 [69]	EEG	CC	Right-hemisphere dominance for phase-locking
Hertrich et al., 2012 [64]	MEG	CC	Phase-locking with right-lateralized peak at 100 ms
Ding and Simon, 2013 [49]	MEG	TRF	Phase-locking at <4 Hz remains stable in noise
Peelle et al., 2013 [32]	MEG	PC	Phase-locking is strongest at 4–7 Hz in intelligible speech
Ding et al., 2014 [113]	MEG	TRF/PC	Phase-locking at 1–4 Hz predicts speech comprehension
Millman et al., 2015 [114]	MEG	CC	Phase-locking at 4–7 Hz regardless of intelligibility
Power et al., 2016 [115]	EEG	SR	Reconstruction of vocoded speech is strongest at 0–2 Hz
Cocktail Party			
Power et al., 2012 [61]	EEG	TRF	Attention elicits left-lateralized peak at 209 ms
Ding and Simon, 2012 [54]	MEG	SR	Attended speech phase-locks at <10 Hz around 100 ms lag
Zion Golumbic et al., 2013 [87]	ECoG	PC	Attended speech phase-locks at 1–7 Hz and 70–150 Hz
Horton et al., 2014 [50]	EEG	CC	Attended phase-locking improves with sample length
O'Sullivan et al., 2015 [83]	EEG	SR	Attended speech encodes maximally at 170–250 ms lag
O'Sullivan et al., 2017 [116]	ECoG	SR	Attention boosts reconstruction accuracy in dynamic switching
Audiovisual Speech			
Crosse et al., 2015 [97]	EEG	SR	AV speech encodes better than A + V at 2–6 Hz
Crosse et al., 2016 [98]	EEG	SR	AV speech improves reconstruction in noise at <3 Hz
Park et al., 2016 [101]	MEG	PC	Cortical activity entrains to lip movements at 1–7 Hz
Linguistic Information			
Di Liberto et al., 2015 [117]	EEG	SR	Cortical activity entrains to phonetic information
Ding et al., 2017 [118]	EEG	PC	Cortical activity entrains to multiple levels concurrently
Falk et al., 2017 [119]	EEG	PC	Phase-locking improves when rhythmic cue precedes speech
Broderick et al., 2018 [120]	EEG	TRF	Neural tracking depends on semantic congruency
Makov et al., 2017 [121]	EEG	PC	Phase-locking at 4 Hz during sleep, but not at higher levels

In summary, neural entrainment to the speech envelope likely reflects, at least in part, prosody perception. Prosodic fluctuations and prosody perception likely contribute to experimental findings linking envelope entrainment to intelligibility, selective attention, and audiovisual integration. Findings discussed in this section are highlighted in Table 1.

5. Developmental and Clinical Relevance of Envelope Entrainment

Children show a reliance on prosody processing from early infancy [122,123]; so, the envelope appears to be a critical tool for early language acquisition. The speech amplitude envelope contributes to the perception of linguistic stress, providing essential information for speech intelligibility and comprehension, e.g., [124]. Infant-directed speech is a manner of speaking that exaggerates prosodic cues, and infants show stronger cortical tracking of the infant-directed speech envelope compared to tracking of adult-directed speech [125]. Individuals who have difficulties with processing cues related to the speech amplitude envelope may demonstrate language-processing deficits [126].

Neuronal oscillatory activity in healthy adults entrains to adult-directed speech at various timescales, e.g., [33]. Frequencies in the delta band range (1–4 Hz) involve slower oscillations and track suprasegmental features of speech, such as phrase patterns, intonation, and stress [33,61]. Prosodic cues are particularly salient in the delta band and may be of particular relevance for envelope entrainment and language acquisition in children. Child-directed speech appears to bolster entrainment at the delta band specifically by amplifying these prosodic features [127]. The accuracy of delta band entrainment may also be indicative of higher-level linguistic abilities, as entrainment at the 0–2 Hz band is positively correlated with literacy [115,128]. The delta band may be crucially important because it provides the foundation for hierarchical linguistic structures of the incoming speech signal [33]. This could, in turn, affect cross-frequency neural synchronization, which may be particularly informative for the development of speech comprehension [129].

Autism spectrum disorders (ASD) are associated with atypical processing of various sensory modalities [130]. Individuals with ASD show less efficient neural integration of audio and visual information in non-speech [131] and speech input [132]. This is related to the temporal binding hypothesis in ASD, which suggests that these individuals have a deficit in synchronization across neural networks [133]. Jochaut et al. [134] showed deficient speech envelope tracking using fMRI and EEG when individuals with ASD perceive congruent audiovisual information. Possible impairment in coupling rhythms into oscillatory hierarchies could contribute to these results [135], and examining language deficits in ASD as oscillopathic traits may be a promising step forward in understanding these disorders [136,137].

Developmental dyslexia is a disorder of reading and spelling difficulties not associated with cognitive deficits or overt neurological conditions, and it is often considered a disorder of phonological processing skills [138]. Dyslexia is believed to affect the temporal coding in the auditory and visual modalities [139,140], and individuals with dyslexia often have difficulty identifying syllable structure or rhyme schemes, see [141]. The speech envelope is important to study in dyslexia because it carries syllable pattern information, and Abrams et al. [142] reported delayed phase-locking to the envelope in individuals with dyslexia. Specifically, the delta band in neuronal oscillations can reveal anomalies such as atypical phase of entrainment [43,143] and poor envelope reconstructions [115], which may ultimately have a downstream effect on establishing phonological representations [25]. Because the delta band reflects prosodic fluctuations, the atypical entrainment in this range suggests that individuals with dyslexia may have impaired encoding at the prosodic linguistic level [61].

Developmental language disorder (DLD) affects language abilities while leaving other cognitive skills intact, and it is sometimes studied in parallel with dyslexia due to similar deficits in phonological and auditory processing [144,145]. The prosodic phrasing hypothesis [146] suggests that children with DLD have difficulty detecting rhythmic patterns in speech, particularly related to impaired sensitivity to amplitude rise time [147] and sound duration [126], and difficulties in processing accelerated speech rate [148]. Given the growing behavioral evidence suggesting that children with DLD have deficits in prosody perception (see [149]), it stands to reason that they would show poor speech envelope entrainment, particularly in the delta frequency band [33]. To our knowledge, there has not been an electrophysiological study looking at neural entrainment to the speech envelope in children with DLD, but this would be an illuminating endeavor.

6. Directions for Future Research

There have been many recent advances related to speech envelope entrainment, and we argue that prosody has had a substantial—though at times underrated—role in many studies. It is well accepted that prosodic cues facilitate speech processing, e.g., [3], and these suprasegmental features are represented in the amplitude envelope, e.g., [64]. Therefore, studies investigating speech envelope entrainment inherently capture a response to prosody to some degree, yet the underlying mechanisms of prosody perception, and their effect on speech processing, remain somewhat a mystery. We suggest that including experimental manipulations of the prosodic dimensions of speech in future studies may inform the findings of previous works, and it may shed light on the future interpretation of entrainment, particularly in the low-frequency range. Ding et al. [118] have shown that removing prosodic cues from speech weakens envelope entrainment, which suggests that synchrony between neural oscillations and the speech envelope reflects perception of the acoustic manifestations of prosody, and future work should continue testing this relationship. More broadly, we present a series of potential future directions in Table 2.

Table 2. Potential future directions including key points and methodological considerations.

Future Directions	
Key Point	**Potential Directions and Methodological Considerations**
Prosodic characteristics of stimuli should be controlled and well-described	• Is it feasible to equalize the prosodic dimension across experimental conditions that are not meant to isolate prosody? At the least, authors could describe the metrical structure of speech stimuli in studies that examine entrainment to envelope features. • What is the variability of neural entrainment to stimuli that differ in prosodic structure?
Role of repetition in establishing neural entrainment to prosodic cues	• What are the implications of hearing the same sentence/stimulus repeated many times versus hearing novel speech? Repetition affects semantic and syntactic expectancies, as well as expectations for the unfolding envelope of the signal. • What is the relationship between predictive neural processes, entrainment to the envelope, and intelligibility? How do these concepts relate to prosody?
Low-frequency envelope fluctuations correlate with the syntactic structure of speech and are relevant to language development	• Does entraining to envelope phrase boundary markers (such as pauses and phrase-final lengthening that correlate with important syntactical information) explain variance in syntactic processing? • How does detection of these cues evolve over the course of childhood language development? • Is neural entrainment to the envelope a potential signature of development of sensitivity to these cues?
Individuals vary in their sensitivity to prosody	• Does neural entrainment to the envelope reflect how individuals differ in their prosodic sensitivity (when measured as a separate behavioral trait)? • Can environmental and genetic factors such as musical training and music aptitude affect individual differences in neural entrainment to speech? • Do some individuals with developmental disabilities have impaired neural entrainment to the envelope? How does this differ among different neurodevelopmental disorders? Is there a causal impact of this impairment on their speech/language/reading development? • Can speech/language/reading therapy enhance sensitivity to prosody via increased neural entrainment to the envelope (as a mediating mechanism to improving speech/language/reading outcomes)?

Synchronization occurs when internal oscillators adjust their phase and period to rate changes of speech rhythm, e.g., [150]. According to the dynamic attending theory [151,152], attentional effort is not uniformly distributed over time, but rather, it occurs periodically with salient sensory input. Prosody offers meaningful information through stressed syllables, which gives attentional rhythms a structure for scaffolding speech processing mechanisms [108]. Suprasegmental elements are present in a wide array of stimuli that demonstrate neural entrainment to the speech envelope (e.g., intelligible and unintelligible speech; attended and unattended speech; audiovisual and audio only and visual only speech). The suprasegmental cues may be one reason why stimuli of varying salience continue to reveal entrainment. Future empirical investigations may consider how prosody supports neural entrainment under these different experimental conditions.

Of course, prosodic fluctuations alone cannot fully explain neural entrainment to speech or speech comprehension [34]. When Ding and colleagues [118,153] removed prosodic cues from connected speech stimuli, they did find some low-frequency entrainment (<10 Hz), which they attribute to syntactic processing. However, they pointed out that neural tracking would likely be more prominent in natural speech with the addition of rich prosodic information. In spoken English, syntax can exist without prosody, but the inclusion of prosody certainly facilitates syntactic processing (with phrase segmentation, pitch inflection, etc.). Therefore, further study of prosodic versus syntactical manipulations will shed light on their respective contributions—and their interaction—to neural entrainment to speech, including when examined together with behavioral measures of speech comprehension. Studies have shown that phonetic [117] and semantic [120] levels of processing also contribute to neural activity at different hierarchical timescales. It may be informative to consider how prosodic cues organize and facilitate processing at these different levels. Future work should attempt to isolate prosodic cues from phonetic and semantic details to specify the contributions of prosody to these other structures in continuous speech. This could be accomplished by restricting prosodic cues (using monotone pitch and constant word durations, as in [118]) or by creating stimuli with a prosodic mismatch (using unpredictable changes in amplitude, pitch, and duration). These manipulations would allow researchers to more directly target the role of prosody in entrainment.

Examining the links between prosody and neural encoding of the speech envelope may also have relevance for additional topics and clinical populations. For example, it has been shown that features of prosody are directly linked to emotional expressiveness in speech [6], and one novel area of research would be to connect patterns of envelope entrainment with perception of emotional states. This would likely have implications for the clinical populations discussed above, as well as typical emotional development. Other recent work has investigated rhythmic cueing and temporal dynamics of speech in patients with Parkinson's disease [154], aphasia [155], and even blindness [156]. Because these populations show difficulty with prosodic cues in speech, a next step could be to examine speech envelope entrainment in these individuals to examine if there is a neural deficit in prosody encoding.

As conveyed in this review, low-frequency neural oscillations likely reflect in part a response to prosodic cues in speech. Future research can investigate how prosody impacts neural envelope entrainment and scaffolds higher-level speech processing, as well as examine individual differences in prosody perception and neural entrainment. Future research in speech entrainment ought to search for connections to prosody perception and determine what it takes to get the speech envelope signed, sealed, and delivered to the cortex.

Funding: The authors were supported by the National Institute on Deafness and Other Communication Disorders and the Office of Behavioral and Social Sciences Research of the National Institutes of Health under Award Numbers R03DC014802, 1R21DC016710-01, and K18DC017383. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. This work was additionally supported by the Program for Music, Mind and Society at Vanderbilt (with funding from the Trans-Institutional Programs Initiative), the VUMC Faculty Research Scholars Program, and the Department of Otolaryngology at Vanderbilt University Medical Center.

Acknowledgments: The authors would like to thank Duane Watson, Cyrille Magne, Stephen Camarata, and anonymous reviewers for invaluable theoretical and conceptual insight to this review, as well as Edmund Lalor, Giovanni Di Liberto, Fleur Bouwer, and Andrew Lotto for thoughtful methodological considerations herein.

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

References

1. Kunert, R.; Jongman, S.R. Entrainment to an auditory signal: Is attention involved? *J. Exp. Psychol. Gen.* **2017**, *146*, 77–88. [CrossRef] [PubMed]
2. Dahan, D.; Tanenhaus, M.K.; Chambers, C.G. Accent and reference resolution in spoken-language comprehension. *J. Mem. Lang.* **2002**, *47*, 292–314. [CrossRef]
3. Pitt, M.A.; Samuel, A.G. The use of rhythm in attending to speech. *J. Exp. Psychol. Hum. Percept. Perform.* **1990**, *16*, 564–573. [CrossRef] [PubMed]
4. Scherer, K.R. Vocal affect expression: A review and a model for future research. *Psychol. Bull.* **1986**, *99*, 143–165. [CrossRef] [PubMed]
5. Zentner, M.; Grandjean, D.; Scherer, K.R. Emotions evoked by the sound of music: Characterization, classification, and measurement. *Emotion* **2008**, *8*, 494–521. [CrossRef] [PubMed]
6. Coutinho, E.; Dibben, N. Psychoacoustic cues to emotion in speech prosody and music. *Cogn. Emot.* **2013**, *27*, 658–684. [CrossRef] [PubMed]
7. Scherer, K.R.; Banse, R.; Wallbott, H.G. Emotion Inferences from Vocal Expression Correlate Across Languages and Cultures. *J. Cross-Cult. Psychol.* **2001**, *32*, 76–92. [CrossRef]
8. Thompson, W.F.; Balkwill, L.-L. Decoding speech prosody in five languages. *Semiotica* **2006**, *158*, 407–424. [CrossRef]
9. Watson, D.; Gibson, E. Intonational phrasing and constituency in language production and comprehension*. *Stud. Linguist.* **2005**, *59*, 279–300. [CrossRef]
10. Liu, F.; Jiang, C.; Wang, B.; Xu, Y.; Patel, A.D. A music perception disorder (congenital amusia) influences speech comprehension. *Neuropsychologia* **2015**, *66*, 111–118. [CrossRef]
11. Magne, C.; Astesano, C.; Aramaki, M.; Ystad, S.; Kronland-Martinet, R.; Besson, M. Influence of Syllabic Lengthening on Semantic Processing in Spoken French: Behavioral and Electrophysiological Evidence. *Cereb. Cortex* **2007**, *17*, 2659–2668. [CrossRef] [PubMed]
12. Gervain, J.; Werker, J.F. Prosody cues word order in 7-month-old bilingual infants. *Nat. Commun.* **2013**, *4*, 1490. [CrossRef]
13. Nazzi, T.; Ramus, F. Perception and acquisition of linguistic rhythm by infants. *Speech Commun.* **2003**, *41*, 233–243. [CrossRef]
14. Soderstrom, M. The prosodic bootstrapping of phrases: Evidence from prelinguistic infants. *J. Mem. Lang.* **2003**, *49*, 249–267. [CrossRef]
15. Shintel, H.; Anderson, N.L.; Fenn, K.M. Talk this way: The effect of prosodically conveyed semantic information on memory for novel words. *J. Exp. Psychol. Gen.* **2014**, *143*, 1437–1442. [CrossRef] [PubMed]
16. Tzeng, C.Y.; Duan, J.; Namy, L.L.; Nygaard, L.C. Prosody in speech as a source of referential information. *Lang. Cogn. Neurosci.* **2018**, *33*, 512–526. [CrossRef]
17. Gordon, R.L.; Shivers, C.M.; Wieland, E.A.; Kotz, S.A.; Yoder, P.J.; Devin McAuley, J. Musical rhythm discrimination explains individual differences in grammar skills in children. *Dev. Sci.* **2015**, *18*, 635–644. [CrossRef] [PubMed]
18. Holt, C.M.; Yuen, I.; Demuth, K. Discourse Strategies and the Production of Prosody by Prelingually Deaf Adolescent Cochlear Implant Users. *Ear Hear.* **2017**, *38*, e101–e108. [CrossRef]
19. Goswami, U.; Gerson, D.; Astruc, L. Amplitude envelope perception, phonology and prosodic sensitivity in children with developmental dyslexia. *Read. Writ.* **2010**, *23*, 995–1019. [CrossRef]
20. Grossman, R.B.; Bemis, R.H.; Plesa Skwerer, D.; Tager-Flusberg, H. Lexical and Affective Prosody in Children With High-Functioning Autism. *J. Speech Lang. Hear. Res.* **2010**, *53*, 778. [CrossRef]
21. Fletcher, J. The Prosody of Speech: Timing and Rhythm. In *The Handbook of Phonetic Sciences*; Hardcastle, W.J., Laver, J., Gibbon, F.E., Eds.; Blackwell Publishing Ltd.: Oxford, UK, 2010; pp. 521–602. ISBN 978-1-4443-1725-1.
22. Lehiste, I. *Suprasegmentals*; M.I.T. Press: Cambridge, MA, USA, 1970; ISBN 978-0-262-12023-4.

23. Fernald, A.; Simon, T. Expanded intonation contours in mothers' speech to newborns. *Dev. Psychol.* **1984**, *20*, 104–113. [CrossRef]
24. Greenberg, S.; Carvey, H.; Hitchcock, L.; Chang, S. Temporal properties of spontaneous speech—A syllable-centric perspective. *J. Phon.* **2003**, *31*, 465–485. [CrossRef]
25. Leong, V.; Goswami, U. Acoustic-Emergent Phonology in the Amplitude Envelope of Child-Directed Speech. *PLoS ONE* **2015**, *10*, e0144411. [CrossRef]
26. Jusczyk, P.W.; Hirsh-Pasek, K.; Kemler Nelson, D.G.; Kennedy, L.J.; Woodward, A.; Piwoz, J. Perception of acoustic correlates of major phrasal units by young infants. *Cognit. Psychol.* **1992**, *24*, 252–293. [CrossRef]
27. Cooper, R.P.; Abraham, J.; Berman, S.; Staska, M. The development of infants' preference for motherese. *Infant Behav. Dev.* **1997**, *20*, 477–488. [CrossRef]
28. Houston, D.M.; Jusczyk, P.W.; Kuijpers, C.; Coolen, R.; Cutler, A. Cross-language word segmentation by 9-month-olds. *Psychon. Bull. Rev.* **2000**, *7*, 504–509. [CrossRef]
29. de Carvalho, A.; Dautriche, I.; Lin, I.; Christophe, A. Phrasal prosody constrains syntactic analysis in toddlers. *Cognition* **2017**, *163*, 67–79. [CrossRef]
30. Sharpe, V.; Fogerty, D.; den Ouden, D.-B. The Role of Fundamental Frequency and Temporal Envelope in Processing Sentences with Temporary Syntactic Ambiguities. *Lang. Speech* **2017**, *60*, 399–426. [CrossRef]
31. Rosen, S. Temporal information in speech: Acoustic, auditory and linguistic aspects. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **1992**, *336*, 367–373.
32. Peelle, J.E.; Gross, J.; Davis, M.H. Phase-Locked Responses to Speech in Human Auditory Cortex are Enhanced During Comprehension. *Cereb. Cortex* **2013**, *23*, 1378–1387. [CrossRef]
33. Giraud, A.-L.; Poeppel, D. Cortical oscillations and speech processing: Emerging computational principles and operations. *Nat. Neurosci.* **2012**, *15*, 511–517. [CrossRef]
34. Obleser, J.; Herrmann, B.; Henry, M.J. Neural Oscillations in Speech: Don't be Enslaved by the Envelope. *Front. Hum. Neurosci.* **2012**, *6*, 250. [CrossRef]
35. Joris, P.X.; Schreiner, C.E.; Rees, A. Neural Processing of Amplitude-Modulated Sounds. *Physiol. Rev.* **2004**, *84*, 541–577. [CrossRef]
36. Fritz, J.; Shamma, S.; Elhilali, M.; Klein, D. Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex. *Nat. Neurosci.* **2003**, *6*, 1216–1223. [CrossRef]
37. Frazier, L.; Carlson, K.; Cliftonjr, C. Prosodic phrasing is central to language comprehension. *Trends Cogn. Sci.* **2006**, *10*, 244–249. [CrossRef]
38. Paraouty, N.; Ewert, S.D.; Wallaert, N.; Lorenzi, C. Interactions between amplitude modulation and frequency modulation processing: Effects of age and hearing loss. *J. Acoust. Soc. Am.* **2016**, *140*, 121–131. [CrossRef]
39. Shannon, R.V.; Zeng, F.G.; Kamath, V.; Wygonski, J.; Ekelid, M. Speech recognition with primarily temporal cues. *Science* **1995**, *270*, 303–304. [CrossRef]
40. Lehiste, I.; Olive, J.P.; Streeter, L.A. Role of duration in disambiguating syntactically ambiguous sentences. *J. Acoust. Soc. Am.* **1976**, *60*, 1199–1202. [CrossRef]
41. Adank, P.; Janse, E. Perceptual learning of time-compressed and natural fast speech. *J. Acoust. Soc. Am.* **2009**, *126*, 2649–2659. [CrossRef]
42. Aubanel, V.; Davis, C.; Kim, J. Exploring the Role of Brain Oscillations in Speech Perception in Noise: Intelligibility of Isochronously Retimed Speech. *Front. Hum. Neurosci.* **2016**, *10*, 430. [CrossRef]
43. Leong, V.; Goswami, U. Assessment of rhythmic entrainment at multiple timescales in dyslexia: Evidence for disruption to syllable timing. *Hear. Res.* **2014**, *308*, 141–161. [CrossRef]
44. Dauer, R.M. Stress-timing and syllable-timing reanalyzed. *J. Phon.* **1983**, *11*, 51–62.
45. Liss, J.M.; LeGendre, S.; Lotto, A.J. Discriminating Dysarthria Type From Envelope Modulation Spectra. *J. Speech Lang. Hear. Res.* **2010**, *53*, 1246. [CrossRef]
46. Drullman, R.; Festen, J.M.; Plomp, R. Effect of temporal envelope smearing on speech reception. *J. Acoust. Soc. Am.* **1994**, *95*, 1053–1064. [CrossRef] [PubMed]
47. Ding, N.; Simon, J.Z. Cortical entrainment to continuous speech: Functional roles and interpretations. *Front. Hum. Neurosci.* **2014**, *8*, 311. [CrossRef]
48. Keitel, A.; Gross, J.; Kayser, C. Perceptually relevant speech tracking in auditory and motor cortex reflects distinct linguistic features. *PLOS Biol.* **2018**, *16*, e2004473. [CrossRef] [PubMed]
49. Ding, N.; Simon, J.Z. Adaptive Temporal Encoding Leads to a Background-Insensitive Cortical Representation of Speech. *J. Neurosci.* **2013**, *33*, 5728–5735. [CrossRef] [PubMed]

50. Horton, C.; Srinivasan, R.; D'Zmura, M. Envelope responses in single-trial EEG indicate attended speaker in a 'cocktail party'. *J. Neural Eng.* **2014**, *11*, 046015. [CrossRef] [PubMed]
51. Shamma, S.A.; Elhilali, M.; Micheyl, C. Temporal coherence and attention in auditory scene analysis. *Trends Neurosci.* **2011**, *34*, 114–123. [CrossRef]
52. Arnal, L.H.; Morillon, B.; Kell, C.A.; Giraud, A.-L. Dual Neural Routing of Visual Facilitation in Speech Processing. *J. Neurosci.* **2009**, *29*, 13445–13453. [CrossRef]
53. Mesgarani, N.; Chang, E.F. Selective cortical representation of attended speaker in multi-talker speech perception. *Nature* **2012**, *485*, 233–236. [CrossRef] [PubMed]
54. Ding, N.; Simon, J.Z. Emergence of neural encoding of auditory objects while listening to competing speakers. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 11854–11859. [CrossRef] [PubMed]
55. Leong, V.; Stone, M.A.; Turner, R.E.; Goswami, U. A role for amplitude modulation phase relationships in speech rhythm perception. *J. Acoust. Soc. Am.* **2014**, *136*, 366–381. [CrossRef]
56. Mai, G.; Minett, J.W.; Wang, W.S.-Y. Delta, theta, beta, and gamma brain oscillations index levels of auditory sentence processing. *NeuroImage* **2016**, *133*, 516–528. [CrossRef]
57. Picton, T.W.; Hillyard, S.A.; Krausz, H.I.; Galambos, R. Human auditory evoked potentials. I: Evaluation of components. *Electroencephalogr. Clin. Neurophysiol.* **1974**, *36*, 179–190. [CrossRef]
58. Schmidt-Kassow, M.; Kotz, S.A. Event-related Brain Potentials Suggest a Late Interaction of Meter and Syntax in the P600. *J. Cogn. Neurosci.* **2009**, *21*, 1693–1708. [CrossRef]
59. Marie, C.; Magne, C.; Besson, M. Musicians and the Metric Structure of Words. *J. Cogn. Neurosci.* **2011**, *23*, 294–305. [CrossRef]
60. Astésano, C.; Besson, M.; Alter, K. Brain potentials during semantic and prosodic processing in French. *Cogn. Brain Res.* **2004**, *18*, 172–184. [CrossRef]
61. Power, A.J.; Foxe, J.J.; Forde, E.-J.; Reilly, R.B.; Lalor, E.C. At what time is the cocktail party? A late locus of selective attention to natural speech: A late locus of attention to natural speech. *Eur. J. Neurosci.* **2012**, *35*, 1497–1503. [CrossRef] [PubMed]
62. Horton, C.; D'Zmura, M.; Srinivasan, R. Suppression of competing speech through entrainment of cortical oscillations. *J. Neurophysiol.* **2013**, *109*, 3082–3093. [CrossRef] [PubMed]
63. Ahissar, E.; Nagarajan, S.; Ahissar, M.; Protopapas, A.; Mahncke, H.; Merzenich, M.M. Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 13367–13372. [CrossRef]
64. Hertrich, I.; Dietrich, S.; Trouvain, J.; Moos, A.; Ackermann, H. Magnetic brain activity phase-locked to the envelope, the syllable onsets, and the fundamental frequency of a perceived speech signal: MEG activity phase-locked to speech. *Psychophysiology* **2012**, *49*, 322–334. [CrossRef] [PubMed]
65. Hickok, G.; Poeppel, D. The cortical organization of speech processing. *Nat. Rev. Neurosci.* **2007**, *8*, 393–402. [CrossRef] [PubMed]
66. Jung, T.-P.; Makeig, S.; Westerfield, M.; Townsend, J.; Courchesne, E.; Sejnowski, T.J. Analysis and visualization of single-trial event-related potentials. *Hum. Brain Mapp.* **2001**, *14*, 166–185. [CrossRef] [PubMed]
67. Poeppel, D. The analysis of speech in different temporal integration windows: Cerebral lateralization as 'asymmetric sampling in time'. *Speech Commun.* **2003**, *41*, 245–255. [CrossRef]
68. Ding, N.; Simon, J.Z. Neural coding of continuous speech in auditory cortex during monaural and dichotic listening. *J. Neurophysiol.* **2012**, *107*, 78–89. [CrossRef] [PubMed]
69. Abrams, D.A.; Nicol, T.; Zecker, S.; Kraus, N. Right-Hemisphere Auditory Cortex Is Dominant for Coding Syllable Patterns in Speech. *J. Neurosci.* **2008**, *28*, 3958–3965. [CrossRef] [PubMed]
70. Aiken, S.J.; Picton, T.W. Human Cortical Responses to the Speech Envelope. *Ear Hear.* **2008**, *29*, 139–157. [CrossRef] [PubMed]
71. Crosse, M.J.; Di Liberto, G.M.; Bednar, A.; Lalor, E.C. The Multivariate Temporal Response Function (mTRF) Toolbox: A MATLAB Toolbox for Relating Neural Signals to Continuous Stimuli. *Front. Hum. Neurosci.* **2016**, *10*. [CrossRef] [PubMed]
72. Marmarelis, V.Z. *Nonlinear Dynamic Modeling of Physiological Systems*; IEEE Press Series in Biomedical Engineering; Wiley-Interscience: Hoboken, NJ, USA, 2004; ISBN 978-0-471-46960-5.
73. Boynton, G.M.; Demb, J.B.; Heeger, D.J. fMRI responses in human V1 correlate with perceived stimulus contrast. *NeuroImage* **1996**, *3*, S265. [CrossRef]

74. Ringach, D.; Shapley, R. Reverse correlation in neurophysiology. *Cogn. Sci.* **2004**, *28*, 147–166. [CrossRef]
75. Lalor, E.C.; Foxe, J.J. Neural responses to uninterrupted natural speech can be extracted with precise temporal resolution. *Eur. J. Neurosci.* **2010**, *31*, 189–193. [CrossRef] [PubMed]
76. Lalor, E.C.; Power, A.J.; Reilly, R.B.; Foxe, J.J. Resolving Precise Temporal Processing Properties of the Auditory System Using Continuous Stimuli. *J. Neurophysiol.* **2009**, *102*, 349–359. [CrossRef] [PubMed]
77. Haufe, S.; Meinecke, F.; Görgen, K.; Dähne, S.; Haynes, J.-D.; Blankertz, B.; Bießmann, F. On the interpretation of weight vectors of linear models in multivariate neuroimaging. *NeuroImage* **2014**, *87*, 96–110. [CrossRef] [PubMed]
78. Yang, X.; Wang, K.; Shamma, S.A. Auditory representations of acoustic signals. *IEEE Trans. Inf. Theory* **1992**, *38*, 824–839. [CrossRef]
79. Theunissen, F.E.; David, S.V.; Singh, N.C.; Hsu, A.; Vinje, W.E.; Gallant, J.L. Estimating spatio-temporal receptive fields of auditory and visual neurons from their responses to natural stimuli. *Netw. Comput. Neural Syst.* **2001**, *12*, 289–316. [CrossRef]
80. Mesgarani, N.; David, S.V.; Fritz, J.B.; Shamma, S.A. Influence of Context and Behavior on Stimulus Reconstruction From Neural Activity in Primary Auditory Cortex. *J. Neurophysiol.* **2009**, *102*, 3329–3339. [CrossRef]
81. Nourski, K.V.; Reale, R.A.; Oya, H.; Kawasaki, H.; Kovach, C.K.; Chen, H.; Howard, M.A.; Brugge, J.F. Temporal Envelope of Time-Compressed Speech Represented in the Human Auditory Cortex. *J. Neurosci.* **2009**, *29*, 15564–15574. [CrossRef]
82. Howard, M.F.; Poeppel, D. Discrimination of Speech Stimuli Based on Neuronal Response Phase Patterns Depends on Acoustics But Not Comprehension. *J. Neurophysiol.* **2010**, *104*, 2500–2511. [CrossRef]
83. O'Sullivan, J.A.; Power, A.J.; Mesgarani, N.; Rajaram, S.; Foxe, J.J.; Shinn-Cunningham, B.G.; Slaney, M.; Shamma, S.A.; Lalor, E.C. Attentional Selection in a Cocktail Party Environment Can Be Decoded from Single-Trial EEG. *Cereb. Cortex* **2015**, *25*, 1697–1706. [CrossRef]
84. Cherry, E.C. Some Experiments on the Recognition of Speech, with One and with Two Ears. *J. Acoust. Soc. Am.* **1953**, *25*, 975–979. [CrossRef]
85. Biesmans, W.; Das, N.; Francart, T.; Bertrand, A. Auditory-Inspired Speech Envelope Extraction Methods for Improved EEG-Based Auditory Attention Detection in a Cocktail Party Scenario. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2017**, *25*, 402–412. [CrossRef]
86. McDermott, J.H. The cocktail party problem. *Curr. Biol.* **2009**, *19*, R1024–R1027. [CrossRef]
87. Zion Golumbic, E.M.; Ding, N.; Bickel, S.; Lakatos, P.; Schevon, C.A.; McKhann, G.M.; Goodman, R.R.; Emerson, R.; Mehta, A.D.; Simon, J.Z.; et al. Mechanisms Underlying Selective Neuronal Tracking of Attended Speech at a "Cocktail Party". *Neuron* **2013**, *77*, 980–991. [CrossRef]
88. Peelle, J.E.; Sommers, M.S. Prediction and constraint in audiovisual speech perception. *Cortex* **2015**, *68*, 169–181. [CrossRef] [PubMed]
89. Erber, N.P. Auditory-Visual Perception of Speech. *J. Speech Hear. Disord.* **1975**, *40*, 481. [CrossRef]
90. Peelle, J.E.; Davis, M.H. Neural Oscillations Carry Speech Rhythm through to Comprehension. *Front. Psychol.* **2012**, *3*. [CrossRef]
91. Chandrasekaran, C.; Ghazanfar, A.A. Different Neural Frequency Bands Integrate Faces and Voices Differently in the Superior Temporal Sulcus. *J. Neurophysiol.* **2009**, *101*, 773–788. [CrossRef]
92. Carlyon, R.P.; Cusack, R.; Foxton, J.M.; Robertson, I.H. Effects of attention and unilateral neglect on auditory stream segregation. *J. Exp. Psychol. Hum. Percept. Perform.* **2001**, *27*, 115–127. [CrossRef]
93. Wagner, P.; Malisz, Z.; Kopp, S. Gesture and speech in interaction: An overview. *Speech Commun.* **2014**, *57*, 209–232. [CrossRef]
94. Loehr, D. Aspects of rhythm in gesture and speech. *Gesture* **2007**, *7*, 179–214. [CrossRef]
95. Krahmer, E.; Swerts, M. The effects of visual beats on prosodic prominence: Acoustic analyses, auditory perception and visual perception. *J. Mem. Lang.* **2007**, *57*, 396–414. [CrossRef]
96. McGurk, H.; MacDonald, J. Hearing lips and seeing voices. *Nature* **1976**, *264*, 746–748. [CrossRef] [PubMed]
97. Crosse, M.J.; Butler, J.S.; Lalor, E.C. Congruent Visual Speech Enhances Cortical Entrainment to Continuous Auditory Speech in Noise-Free Conditions. *J. Neurosci.* **2015**, *35*, 14195–14204. [CrossRef] [PubMed]
98. Crosse, M.J.; Di Liberto, G.M.; Lalor, E.C. Eye Can Hear Clearly Now: Inverse Effectiveness in Natural Audiovisual Speech Processing Relies on Long-Term Crossmodal Temporal Integration. *J. Neurosci.* **2016**, *36*, 9888–9895. [CrossRef]

99. Sumby, W.H.; Pollack, I. Visual Contribution to Speech Intelligibility in Noise. *J. Acoust. Soc. Am.* **1954**, *26*, 212–215. [CrossRef]
100. Ross, L.A.; Saint-Amour, D.; Leavitt, V.M.; Molholm, S.; Javitt, D.C.; Foxe, J.J. Impaired multisensory processing in schizophrenia: Deficits in the visual enhancement of speech comprehension under noisy environmental conditions. *Schizophr. Res.* **2007**, *97*, 173–183. [CrossRef]
101. Park, H.; Kayser, C.; Thut, G.; Gross, J. Lip movements entrain the observers' low-frequency brain oscillations to facilitate speech intelligibility. *eLife* **2016**, *5*. [CrossRef] [PubMed]
102. Tye-Murray, N.; Sommers, M.; Spehar, B. Auditory and Visual Lexical Neighborhoods in Audiovisual Speech Perception. *Trends Amplif.* **2007**, *11*, 233–241. [CrossRef]
103. Kösem, A.; Bosker, H.R.; Takashima, A.; Meyer, A.; Jensen, O.; Hagoort, P. Neural Entrainment Determines the Words We Hear. *Curr. Biol.* **2018**, *28*, 2867–2875. [CrossRef]
104. Jäncke, L. The Relationship between Music and Language. *Front. Psychol.* **2012**, *3*. [CrossRef]
105. Hausen, M.; Torppa, R.; Salmela, V.R.; Vainio, M.; Särkämö, T. Music and speech prosody: A common rhythm. *Front. Psychol.* **2013**, *4*. [CrossRef]
106. Bonacina, S.; Krizman, J.; White-Schwoch, T.; Kraus, N. Clapping in time parallels literacy and calls upon overlapping neural mechanisms in early readers: Clapping in time parallels literacy. *Ann. N. Y. Acad. Sci.* **2018**, *1423*, 338–348. [CrossRef]
107. Ozernov-Palchik, O.; Wolf, M.; Patel, A.D. Relationships between early literacy and nonlinguistic rhythmic processes in kindergarteners. *J. Exp. Child Psychol.* **2018**, *167*, 354–368. [CrossRef]
108. Tierney, A.; Kraus, N. Auditory-motor entrainment and phonological skills: Precise auditory timing hypothesis (PATH). *Front. Hum. Neurosci.* **2014**, *8*, 949. [CrossRef]
109. Woodruff Carr, K.; White-Schwoch, T.; Tierney, A.T.; Strait, D.L.; Kraus, N. Beat synchronization predicts neural speech encoding and reading readiness in preschoolers. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 14559–14564. [CrossRef]
110. Doelling, K.B.; Poeppel, D. Cortical entrainment to music and its modulation by expertise. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, E6233–E6242. [CrossRef]
111. Harding, E.E.; Sammler, D.; Henry, M.J.; Large, E.W.; Kotz, S.A. Cortical tracking of rhythm in music and speech. *NeuroImage* **2019**, *185*, 96–101. [CrossRef]
112. Luo, H.; Poeppel, D. Phase Patterns of Neuronal Responses Reliably Discriminate Speech in Human Auditory Cortex. *Neuron* **2007**, *54*, 1001–1010. [CrossRef] [PubMed]
113. Ding, N.; Chatterjee, M.; Simon, J.Z. Robust cortical entrainment to the speech envelope relies on the spectro-temporal fine structure. *NeuroImage* **2014**, *88*, 41–46. [CrossRef]
114. Millman, R.E.; Johnson, S.R.; Prendergast, G. The Role of Phase-locking to the Temporal Envelope of Speech in Auditory Perception and Speech Intelligibility. *J. Cogn. Neurosci.* **2015**, *27*, 533–545. [CrossRef] [PubMed]
115. Power, A.J.; Colling, L.J.; Mead, N.; Barnes, L.; Goswami, U. Neural encoding of the speech envelope by children with developmental dyslexia. *Brain Lang.* **2016**, *160*, 1–10. [CrossRef] [PubMed]
116. O'Sullivan, J.; Chen, Z.; Herrero, J.; McKhann, G.M.; Sheth, S.A.; Mehta, A.D.; Mesgarani, N. Neural decoding of attentional selection in multi-speaker environments without access to clean sources. *J. Neural Eng.* **2017**, *14*, 056001. [CrossRef]
117. Di Liberto, G.M.; O'Sullivan, J.A.; Lalor, E.C. Low-Frequency Cortical Entrainment to Speech Reflects Phoneme-Level Processing. *Curr. Biol.* **2015**, *25*, 2457–2465. [CrossRef] [PubMed]
118. Ding, N.; Melloni, L.; Yang, A.; Wang, Y.; Zhang, W.; Poeppel, D. Characterizing Neural Entrainment to Hierarchical Linguistic Units using Electroencephalography (EEG). *Front. Hum. Neurosci.* **2017**, *11*, 481. [CrossRef] [PubMed]
119. Falk, S.; Lanzilotti, C.; Schön, D. Tuning Neural Phase Entrainment to Speech. *J. Cogn. Neurosci.* **2017**, *29*, 1378–1389. [CrossRef]
120. Broderick, M.P.; Anderson, A.J.; Di Liberto, G.M.; Crosse, M.J.; Lalor, E.C. Electrophysiological Correlates of Semantic Dissimilarity Reflect the Comprehension of Natural, Narrative Speech. *Curr. Biol.* **2018**, *28*, 803–809. [CrossRef] [PubMed]
121. Makov, S.; Sharon, O.; Ding, N.; Ben-Shachar, M.; Nir, Y.; Zion Golumbic, E. Sleep Disrupts High-Level Speech Parsing Despite Significant Basic Auditory Processing. *J. Neurosci.* **2017**, *37*, 7772–7781. [CrossRef]
122. Curtin, S.; Mintz, T.H.; Christiansen, M.H. Stress changes the representational landscape: Evidence from word segmentation. *Cognition* **2005**, *96*, 233–262. [CrossRef]

123. Mehler, J.; Jusczyk, P.; Lambertz, G.; Halsted, N.; Bertoncini, J.; Amiel-Tison, C. A precursor of language acquisition in young infants. *Cognition* **1988**, *29*, 143–178. [CrossRef]
124. Ghitza, O. On the Role of Theta-Driven Syllabic Parsing in Decoding Speech: Intelligibility of Speech with a Manipulated Modulation Spectrum. *Front. Psychol.* **2012**, *3*, 238. [CrossRef] [PubMed]
125. Kalashnikova, M.; Peter, V.; Di Liberto, G.M.; Lalor, E.C.; Burnham, D. Infant-directed speech facilitates seven-month-old infants' cortical tracking of speech. *Sci. Rep.* **2018**, *8*, 13745. [CrossRef] [PubMed]
126. Richards, S.; Goswami, U. Auditory Processing in Specific Language Impairment (SLI): Relations With the Perception of Lexical and Phrasal Stress. *J. Speech Lang. Hear. Res.* **2015**, *58*, 1292. [CrossRef] [PubMed]
127. Leong, V.; Kalashnikova, M.; Burnham, D.; Goswami, U. The Temporal Modulation Structure of Infant-Directed Speech. *Open Mind* **2017**, *1*, 78–90. [CrossRef]
128. Molinaro, N.; Lizarazu, M.; Lallier, M.; Bourguignon, M.; Carreiras, M. Out-of-synchrony speech entrainment in developmental dyslexia: Altered Cortical Speech Tracking in Dyslexia. *Hum. Brain Mapp.* **2016**, *37*, 2767–2783. [CrossRef] [PubMed]
129. Leong, V.; Goswami, U. Impaired extraction of speech rhythm from temporal modulation patterns in speech in developmental dyslexia. *Front. Hum. Neurosci.* **2014**, *8*, 96. [CrossRef]
130. Marco, E.J.; Hinkley, L.B.N.; Hill, S.S.; Nagarajan, S.S. Sensory Processing in Autism: A Review of Neurophysiologic Findings. *Pediatr. Res.* **2011**, *69*, 48R–54R. [CrossRef]
131. Brandwein, A.B.; Foxe, J.J.; Butler, J.S.; Russo, N.N.; Altschuler, T.S.; Gomes, H.; Molholm, S. The Development of Multisensory Integration in High-Functioning Autism: High-Density Electrical Mapping and Psychophysical Measures Reveal Impairments in the Processing of Audiovisual Inputs. *Cereb. Cortex* **2013**, *23*, 1329–1341. [CrossRef] [PubMed]
132. Stevenson, R.A.; Siemann, J.K.; Schneider, B.C.; Eberly, H.E.; Woynaroski, T.G.; Camarata, S.M.; Wallace, M.T. Multisensory Temporal Integration in Autism Spectrum Disorders. *J. Neurosci.* **2014**, *34*, 691–697. [CrossRef]
133. Brock, J.; Brown, C.C.; Boucher, J.; Rippon, G. The temporal binding deficit hypothesis of autism. *Dev. Psychopathol.* **2002**, *14*, 209–224. [CrossRef]
134. Jochaut, D.; Lehongre, K.; Saitovitch, A.; Devauchelle, A.-D.; Olasagasti, I.; Chabane, N.; Zilbovicius, M.; Giraud, A.-L. Atypical coordination of cortical oscillations in response to speech in autism. *Front. Hum. Neurosci.* **2015**, *9*, 171. [CrossRef] [PubMed]
135. Kikuchi, M.; Yoshimura, Y.; Hiraishi, H.; Munesue, T.; Hashimoto, T.; Tsubokawa, T.; Takahashi, T.; Suzuki, M.; Higashida, H.; Minabe, Y. Reduced long-range functional connectivity in young children with autism spectrum disorder. *Soc. Cogn. Affect. Neurosci.* **2015**, *10*, 248–254. [CrossRef] [PubMed]
136. Benítez-Burraco, A.; Murphy, E. The Oscillopathic Nature of Language Deficits in Autism: From Genes to Language Evolution. *Front. Hum. Neurosci.* **2016**, *10*, 120. [CrossRef] [PubMed]
137. Simon, D.M.; Wallace, M.T. Dysfunction of sensory oscillations in Autism Spectrum Disorder. *Neurosci. Biobehav. Rev.* **2016**, *68*, 848–861. [CrossRef] [PubMed]
138. Stanovich, K.E. Refining the Phonological Core Deficit Model. *Child Psychol. Psychiatry Rev.* **1998**, *3*, 17–21. [CrossRef]
139. Lallier, M.; Thierry, G.; Tainturier, M.-J.; Donnadieu, S.; Peyrin, C.; Billard, C.; Valdois, S. Auditory and visual stream segregation in children and adults: An assessment of the amodality assumption of the 'sluggish attentional shifting' theory of dyslexia. *Brain Res.* **2009**, *1302*, 132–147. [CrossRef] [PubMed]
140. Goswami, U. A temporal sampling framework for developmental dyslexia. *Trends Cogn. Sci.* **2011**, *15*, 3–10. [CrossRef] [PubMed]
141. Ziegler, J.C.; Goswami, U. Reading Acquisition, Developmental Dyslexia, and Skilled Reading Across Languages: A Psycholinguistic Grain Size Theory. *Psychol. Bull.* **2005**, *131*, 3–29. [CrossRef] [PubMed]
142. Abrams, D.A.; Nicol, T.; Zecker, S.; Kraus, N. Abnormal Cortical Processing of the Syllable Rate of Speech in Poor Readers. *J. Neurosci.* **2009**, *29*, 7686–7693. [CrossRef]
143. Power, A.J.; Mead, N.; Barnes, L.; Goswami, U. Neural entrainment to rhythmic speech in children with developmental dyslexia. *Front. Hum. Neurosci.* **2013**, *7*, 777. [CrossRef]
144. Goswami, U.; Cumming, R.; Chait, M.; Huss, M.; Mead, N.; Wilson, A.M.; Barnes, L.; Fosker, T. Perception of Filtered Speech by Children with Developmental Dyslexia and Children with Specific Language Impairments. *Front. Psychol.* **2016**, *7*, 791. [CrossRef]

145. Przybylski, L.; Bedoin, N.; Krifi-Papoz, S.; Herbillon, V.; Roch, D.; Léculier, L.; Kotz, S.A.; Tillmann, B. Rhythmic auditory stimulation influences syntactic processing in children with developmental language disorders. *Neuropsychology* **2013**, *27*, 121–131. [CrossRef]
146. Cumming, R.; Wilson, A.; Leong, V.; Colling, L.J.; Goswami, U. Awareness of Rhythm Patterns in Speech and Music in Children with Specific Language Impairments. *Front. Hum. Neurosci.* **2015**, *9*, 672. [CrossRef]
147. Beattie, R.L.; Manis, F.R. Rise Time Perception in Children With Reading and Combined Reading and Language Difficulties. *J. Learn. Disabil.* **2013**, *46*, 200–209. [CrossRef]
148. Guiraud, H.; Bedoin, N.; Krifi-Papoz, S.; Herbillon, V.; Caillot-Bascoul, A.; Gonzalez-Monge, S.; Boulenger, V. Don't speak too fast! Processing of fast rate speech in children with specific language impairment. *PLoS ONE* **2018**, *13*, e0191808. [CrossRef]
149. Cumming, R.; Wilson, A.; Goswami, U. Basic auditory processing and sensitivity to prosodic structure in children with specific language impairments: A new look at a perceptual hypothesis. *Front. Psychol.* **2015**, *6*, 972. [CrossRef]
150. Kotz, S.A.; Schwartze, M.; Schmidt-Kassow, M. Non-motor basal ganglia functions: A review and proposal for a model of sensory predictability in auditory language perception. *Cortex* **2009**, *45*, 982–990. [CrossRef]
151. Large, E.W.; Jones, M.R. The dynamics of attending: How people track time-varying events. *Psychol. Rev.* **1999**, *106*, 119–159. [CrossRef]
152. Iversen, J.R.; Repp, B.H.; Patel, A.D. Top-Down Control of Rhythm Perception Modulates Early Auditory Responses. *Ann. N. Y. Acad. Sci.* **2009**, *1169*, 58–73. [CrossRef]
153. Ding, N.; Melloni, L.; Zhang, H.; Tian, X.; Poeppel, D. Cortical tracking of hierarchical linguistic structures in connected speech. *Nat. Neurosci.* **2015**, *19*, 158. [CrossRef]
154. Kotz, S.A.; Gunter, T.C. Can rhythmic auditory cuing remediate language-related deficits in Parkinson's disease?: Rhythmic auditory cuing and language. *Ann. N. Y. Acad. Sci.* **2015**, *1337*, 62–68. [CrossRef]
155. Knilans, J.; DeDe, G. Online Sentence Reading in People With Aphasia: Evidence From Eye Tracking. *Am. J. Speech Lang. Pathol.* **2015**, *24*, S961. [CrossRef]
156. Van Ackeren, M.J.; Barbero, F.M.; Mattioni, S.; Bottini, R.; Collignon, O. Neuronal populations in the occipital cortex of the blind synchronize to the temporal dynamics of speech. *eLife* **2018**, *7*, e31640. [CrossRef]