ADVANCED REVIEW

WIREs COGNITIVE SCIENCE WILEY

Syllable as a unit of information transfer in linguistic communication: The entropy syllable parsing model

Evie A. Malaia[1] | Ronnie B. Wilbur[2,3]

[1]Department of Communicative Disorders, University of Alabama, Tuscaloosa, Alabama

[2]Department of Speech, Language, Hearing Sciences, College of Health and Human Sciences, Purdue University, West Lafayette, Indiana

[3]Linguistics, School of Interdisciplinary Studies, College of Liberal Arts, Purdue University, West Lafayette, Indiana

Correspondence
Evie A. Malaia, Department of Communicative Disorders, University of Alabama, Tuscaloosa, AL 35487.
Email: eamalaia@ua.edu

Abstract

To understand human language—both spoken and signed—the listener or viewer has to parse the continuous external signal into components. The question of what those components are (e.g., phrases, words, sounds, phonemes?) has been a subject of long-standing debate. We re-frame this question to ask: What properties of the incoming visual or auditory signal are indispensable to eliciting language comprehension? In this review, we assess the phenomenon of language parsing from modality-independent viewpoint. We show that the interplay between dynamic changes in the entropy of the signal and between neural entrainment to the signal at syllable level (4–5 Hz range) is causally related to language comprehension in both speech and sign language. This modality-independent Entropy Syllable Parsing model for the linguistic signal offers insight into the mechanisms of language processing, suggesting common neurocomputational bases for syllables in speech and sign language.

This article is categorized under:

Linguistics > Linguistic Theory
Linguistics > Language in Mind and Brain
Linguistics > Computational Models of Language
Psychology > Language

KEYWORDS

entrainment, parsing, segmentation, sign language, syllable

1 | INTRODUCTION

Syllables appear to be everywhere. They feature prominently in tip-of-the-tongue phenomena, meter and rhyme in poetry and music, hyphenation of words in written texts, clues in charades and crossword puzzles, assignment of linguistic stress, production of language games like Pig Latin, and determination of melodious names for commercial products. Yet, the syllable is one of the most elusive of all linguistic notions. Roger Lass, with tongue in cheek, quipped: "Everyone knows that 'a syllable' is what syllable has three of" (Lass, 1984).

As understanding of sign languages as fully fledged human languages became prevalent, the question of a modality-independent definition of syllable became a major puzzle for linguistics. From the first presentation on sign language syllables by Nancy Chinchor in 1978, until the publication of Coulter's edited volume Current issues in ASL phonology in 1993, many researchers have addressed the question of whether sign languages had syllables and, if so, how they were constructed (Allen, Wilbur, & Schick, 1991; Corina, 1990; Liddell & Johnson, 1989; Padden & Perlmutter, 1987; Sandler, 1989; Wilbur & Allen,

1991; Wilbur & Nolen, 1986; Wilbur & Schick, 1987). None of those papers questioned the existence of sign language syllables, but rather focused on evidence for the functions that syllables performed in phonological and morphological processes, and, more controversially, what the internal structure of a signed syllable looked like, especially whether there were segments comparable to consonants and vowels for speech. At about the same time, the discussion among spoken language phonologists and phoneticians was heading in the opposite direction: away from the definition of syllable as sequences of segments. Goldstein and Browman (1986) provided evidence that a syllable corresponds to a single gesture on the laryngeal tier, beginning what has come to be known as the Articulatory Phonology approach. While these nonsegmental syllable definitions for signing and speech moved closer toward each other, defining units of comparable durations and similar linguistic behaviors (Wilbur, 2010), there still appears to be a gap in the cross-modal syllable definition, as, rather obviously, sign production does not require larynx movement.

In this review, we turn our attention not to how the syllable is produced, but to how it is represented in the signal and how it is parsed and extracted from the signal by the human brain. Combining recent achievements in neurolinguistic analysis of speech and sign, we arrive at a modality-independent definition of the syllable structure as the rise and fall of signal entropy, which drives neural entrainment at the ~4 Hz frequency. As we show, this approach is well-supported by recent neurolinguistic data, applicable for both spoken and signed languages, and allows for quantitative analysis of human language perception and production in neurotypical and atypical populations.

The use of information entropy as a measure of the communicative signal dates back to Shannon (1948). Generally, entropy is a measure of newness of the incoming signal, or uncertainty over the next item coming forth in a sequence. In speech, one can think of entropy as the probability of hearing the same or a different phoneme in the next time slot (if the probability of a repeat is low, the entropy, or the amount of information that is possible to transfer within a unit of time, is high). In sign language, the equivalent measure of entropy is the amount of change in the visual field of the signer, including motion of all articulators—hands, face, head, and body. In mathematical terms, information entropy is the rate (over time) at which information is produced by any source of data. The amount of information being transferred can be calculated to any modality (visual, auditory), or a specific bandwidth within the modality (e.g., the frequency spectra of human hearing or vision). Entropy, however, should not be confused with the spectrogram of the signal—rather, it is a representation of the spectral change in the signal in question. Unlike the spectrum, which is represented over time, entropy is a metric of the use frequency in the signal.

2 | SYLLABLES IN LANGUAGE AND LINGUISTICS

Historically, as noted by Hyman (1975), three questions recurred in discussions of syllables:

1. How does one define the syllable?
2. How does one determine syllable boundaries?
3. Is the syllable a necessary concept?

This last question has a contentious history. Various syllable replacement attempts were made, which focused on individual segments constituting the syllable (such as consonants and vowels), or parts of segments, such as distinctive features (Chomsky & Halle, 1968). Yet, Chomsky and Halle do not define or discuss the notion of syllable before invoking the notion of syllabic peak as the definition for their feature [syllabic]. Syllables were the focus of much phonological research in the 1970s (Hooper, 1972; Kahn, 1976; Vennemann, 1972), when various models were proposed in generative, metrical, and autosegmental phonology, with the primary issue being whether syllables are flat or have internal hierarchical structure.

Clements and Keyser (1983) proposed stating syllable generalization in terms of abstract units C and V, which do not necessarily coincide with notions of consonant and vowel. In a variant suggestion, Hyman (1985) proposed that what really matters is whether a segment carries phonological weight; he would thus eliminate the Consonant Vowel tier and replace it with a phonological weight tier. Hyman also argued that syllables are not absolute linguistic universals, finding no justification for positing syllables in the eastern Nigerian language Gokana. While speakers of Gokana can count syllables or provide other psycholinguistic evidence of syllable awareness in perception and production, there are simply no rules or constraints in Gokana phonology that refer to the syllable as a unit (e.g., no constraints on consonants to specific positions in the syllable or tonal patterns that can only occur on syllables). What Hyman's argument highlights is the dual nature of syllables: (a) the linguistic unit that may be relevant to the statement of phonological generalizations and (b) the psycholinguistic unit that can be quantified. Lass (1984) refers to these as the phonological syllable and the phonetic syllable, respectively. It is also useful to

consider how the phonetic syllable and the phonological syllable can coexist—indeed, how both may be necessary. Hyman (1975) uses the word "bedroom" to demonstrate how these two types of syllables may coexist but have different boundaries. Phonologically and morphologically, "bedroom" has two syllables: /bɛd.rum/. Phonetically "bedroom" is pronounced [bɛ.ʤrum] because the [d] is adjacent to the [r], hence the intermediate [bɛ.drum]. This distinction illustrates that acceptance and use of a phonological model of syllables does not preclude additional syllable analysis at the phonetic level, or vice versa. Indeed, resyllabification creates numerous situations in which the syllabification associated with a lexical entry and additional affixes is different from the syllabification after all relevant associations, deletions, and changes have been effected (Clements & Keyser, 1983; Goldsmith, 1979). As one can see from this discussion, syllables are not a necessary concept at the level of phonological description, but they are presumed to always be present at least at the phonetic level.

Other proposed syllable definitions used articulation, acoustics, and phonological characteristics as bases. One approach to defining the phonetic syllable is to specify articulatory correlates—treating syllables as the result of what speakers produce. Ladefoged (1982) considered and rejected two such proposals: (a) each syllable is initiated by a chest pulse (the theory originated by Stetson, 1951), a contraction of the rib cage that forces more air out of the lungs; and (b) a syllable can be defined in terms of some combination of laryngeal and respiratory activity. Further progress toward an articulatorily based definition was that of Browman and Goldstein (1986), building on Anderson's (1974) observations that articulation can be decomposed into four subsystems: an energy source, a laryngeal system, an oral system, and a nasal system. Syllables result from the coordinated trajectories of articulatory gestures in each of these systems, with the laryngeal tier being in essence the syllable "spine."

Acoustic approaches to syllable research focused on the properties of the speech stream: What the listener hears. The best-known acoustic definition of syllable is based on sonority—defined as the loudness of a sound relative to other sounds that have the same length, stress, and pitch. Each sound in language has its own inherent sonority, with vowels more sonorant than consonants and with variations in each group; the ranking is known as the sonority hierarchy. The vowel /a/ is more sonorant than /i/, and nasal consonants are more sonorant than plain stops. Syllable peak (nucleus) requires presence of a phoneme that is more sonorant than phonemes in the (optional) onset and coda. Ladefoged (1982) observed that this approach does not account for all the facts of language (such as words like "spa" that have one syllable but two sonority peaks). Other problems with this definition include existence of phrase pairs, such as *hidden aims* and *hid names*, which have the same number of sonority peaks, but a different number of syllables (three in the first, two in the second).

Without the benefit of a phonetic definition, abstract models of syllable structure have been used for at least two thousand years, deriving from early attempts to describe rhyming in poetry. In rhyme, there is usually agreement in the terminal sounds, including the last stressed vowel and any sounds following it, while the preceding sounds are different. The sounds that agree are called the "rhyme." Early linguistic efforts to describe syllables divided them into two parts, the rhyme and what has come to be known as the onset. Another term for the component of the syllable, the nucleus, refers to the vowel and any following semivowel as a unit. Syllable peak usually refers to the vowel itself, while the semivowel (if present) is frequently referred to as the satellite. Consonants following the peak vowel and satellite, or the nucleus, are referred to as the coda. These pieces can be combined in a variety of ways. A flat model of the syllable might have the onset, peak or nucleus, and coda as ordered single-level units. Hyman (1975) used three parts of a syllable (onset, peak, and coda) but argued that the only phonologically relevant division was that between the onset and the rhyme (his "core"). In this type of model, the syllable has an internal hierarchical structure, with the division between onset and rhyme being more important than that between the peak and the coda (see also Cairns & Feinstein, 1982; Clements & Keyser, 1983; Goldsmith, 1979; Steriade, 1988). Yet another family of approaches to syllable definition focused on the syllable boundaries as opposed to syllable content. Kahn (1976) proposed a rule-based procedure for syllabification without proposing any model of the syllable internal makeup. Kenstowicz and Kisseberth (1979) accepted Kahn's rules as providing a "consistent definition of the syllable that is required for the statement of several other phonological rules of English", in accordance with the arguments made by Clements and Keyser that additional internal structure for the syllable was unnecessary.

Initial studies of sign language made no distinction between sign and syllable, in large part because there is a strong trend for lexical signs to be monosyllabic. Decomposition of a sign identified a "simultaneous" bundle of components, the so-called "big four": handshape, location, movement, and orientation (Battison, 1978; Bellugi & Klima, 1979; Friedman, 1974, 1976; Siple, 1978; Stokoe, 1960; Wilbur, 1979). Some studies at that time did suggest that there are syllables in ASL, and that their internal structure was also sequential (Chinchor, 1981; Kegl & Wilbur, 1976; Liddell, 1984; Newkirk, 1979; Newkirk, 1998). Wilbur (2010) provides a more elaborate explanation of the difference between a sign and a syllable, but for purposes here, it suffices to note that to produce a sign there must be at least one syllable; there are disyllabic lexical signs and reduplicated signs with two, three, or somewhat more syllables; and finally that morphemes such as classifier handshapes and motion verbs

can be themselves smaller than a syllable, meaning that they are bound morphemes and must combine with other morphemes to compose a full syllable.[1] Thus, a sign can be monomorphemic or multimorphemic.

As with spoken syllables, there have been debates about the signed syllable, but the issues have been somewhat different. The field is unanimous with respect to the notion of movement as the syllabic core; but there, unanimity ends. There have been several attempts at defining segmental-sequential syllables, with two types of segments. In each, movement (M) is one type of segment, with Holds (H) (Liddell, 1984) or Locations (L) (Sandler, 1986, 1989) as the other type, yielding syllable types of M, MH (or ML), HMH (or LML), and so forth. These models make predictions that are falsified by backwards signing (Wilbur & Petersen, 1997). Incorrect predictions from segmental models are more obvious with the verb sign FLY, which would be represented as a single M segment. The predicted backwards version should be the same as the original, because there is nothing available to exchange. Backwards signing shows that the direction of movement of FLY is reversed—if the signer makes the original with an upward movement, fingertips leading, then the reversed form is made with a downward movement, with the wrist end of the hand leading and fingertips trailing. In backwards signing, movements are consistently reversed by exchanging end specification with start specification, as though initial and final features are exchanged on their own tiers: end location with beginning location; end handshape with beginning handshape; end orientation with beginning orientation. In short, a distinctive middle is missing. This view is further supported by a tapping study which showed that, unlike tapping to speech which attracts taps to the consonant burst-vowel onset transition, tapping to signed syllables is equally/randomly distributed throughout the syllable duration, with no apparent tap attractors even with visible events such as contact between the hands or holds following the end of syllables (Wilbur & Allen, 1991). In the syllable-model competition, the model that has survived and is supported by the psycholinguistic research is that proposed by Brentari (1998). The model itself is quite detailed, thus we present only the minimum necessary to further the discussion of syllable similarity across modalities. Brentari takes features as basic building blocks, and divides the syllable into two major featural types, the Inherent Features, which include those features that do not change during production, and the Prosodic Features, which are those features that describe the movement. Each prosodic branch has two timing slots, not three as would be predicted by the segmental models. If a syllable has a handshape change as its primary movement, the movement is nearly always an aperture change (distance between fingers) or open-to-closed or closed-to-open. The first timing slot of the prosodic branch would have the initial setting (say, open) and the second slot would have the final setting (which would be closed, if all fingers are involved). The actual handshape, orientation, and location of the production would be specified in the Inherent Features. In many, but not all, movements, the movement emerges from the difference between the features on the two timing slots. Thus, reversing the feature specifications (in backward signing) results in a different movement than the unreversed form. The Brentari model is consistent with the backwards signing results. Movements that are not the emergent results of between-feature specifications generally have different features, such as specified path shape (circle, arc, square, irregular, which Brentari terms "tracing"). These movements caused particular difficulty for signers when they were asked to sign them backwards, because it is not obvious how to reverse, for example, a circle or wiggling fingers (Wilbur & Petersen, 1997). Thus, the signed syllable debate has been one of syllable-internal structure rather than syllable definition. The two timing-slot syllable model fits well with the general syllable model we discuss below.

Once the need for the concept of the syllable, and the difficulties of defining it separately in auditory and visual domains are put together, it is clear that the concept of the syllable has to be abstract in terms of both modality (not based on articulatory or acoustic properties), and apply cross-linguistically (i.e., on a wide variety of hierarchically organized segmental and suprasegmental features, such as tone or motion) to both production and perception. We apply the information-theory concept of entropy to propose a model of syllable that incorporates neural, visual–auditory, and developmental perspectives, and abstracts the syllable from the articulatory and acoustic properties to the modality-independent level: the Entropy Syllable Parsing model.

3 | PROCESSING OF SIGNAL ENTROPY IN SPEECH AND SIGN

There are competing theoretical viewpoints on the question of relevance, on one hand, of the properties of the signal which might be identified as informative at the neural level (c.f., Malaia, Borneman, & Wilbur, 2016), and the top-down processes required for comprehension, such as attention to the immediate incoming signal (Ding et al., 2018), and long-term linguistic experience (Lidji, Palmer, Peretz, & Morningstar, 2011). We will review experimental evidence for both bottom-up and top-down processing phenomena in the visual and auditory modalities, connecting the pieces of evidence toward a general conclusion—that the parsing (and, as a result, comprehension) of linguistic input is undergirded by the presence of the ~4 Hz

envelope structure in the physical (visual or auditory) signal. This physical rise-and-fall sequence in the signal entropy is commonly understood as "syllable."

Psychology has firmly established that segmentation of reality into separate events is the basis of everyday cognition. Infants learn to break down the continuous stream of sensory input from their environment into separate segments (events) in order to parse and interpret dynamic actions (Baldwin, Baird, Saylor, & Clark, 2001; Loucks & Baldwin, 2009; Olofson & Baldwin, 2011). As adults, we use the ability to segment continuous external input for remembering what happened in the past (Sargent et al., 2013; Swallow, Zacks, & Abrams, 2009), create conceptual representations of our environment (Zacks & Tversky, 2001), and predict and understand sentences and narratives as they unfold in communication (Malaia, Talavage, & Wilbur, 2014; Malaia, Wilbur, & Weber-Fox, 2009; Speer, Zacks, & Reynolds, 2007). The processes involved in parsing the constantly changing external signal can be either bottom-up—i.e., based on parameters of the input, such as velocity of motion in visual scenes (Malaia & Wilbur, 2012; Malaia, Wilbur, & Weber-Fox, 2013; Strickland et al., 2015; Zacks et al., 2001; Zacks, Swallow, Vettel, & McAvoy, 2006), or top-down, that is, based on the pre-existing conceptual model of reality, such as actors' intentions (Zacks, Kumar, Abrams, & Mehta, 2009).

Once an event is identified, it is put in the context of other events—that is, entered into the hierarchy for conceptual processing. For example, one large-scale (coarse) event (putting together a Lego toy) will consist of multiple fine-grained events, such as "review instructions," "set up pieces for each step," "connect pieces according to step 1 of instructions," and so forth. To understand linguistic communication in light of signal parsing, one has to identify the parameters of segmentable chunks of speech and sign language, which contribute critically to intelligibility of the message.

3.1 | Bottom-up processing of spectro-temporal content of the syllable

The two parameters that are critical to understanding parseable components of the continuous signal are the temporal resolution and the amount of change present in the signal within the given time window. In speech science, the key contribution of the low-frequency envelope to intelligibility is undisputed, even if the exact nature of the relevant dimension of the envelope to perception and comprehension is still under discussion. At various points in time, the proposed measures of communicative potential in speech included signal intensity (Oxenham, Boucher, & Kreft, 2017; Shu, Feng, & Chen, 2016), perception-scaled entropy (cochlear-scaled entropy—Kluender & Kiefte, 2006; Stilp & Kluender, 2010, 2016), and Shannon information (mathematical variability) in the signal (Kluender, Stilp, & Lucas, 2019; Lewicki, 2010). The common points among these are, first, the rise-and-fall sequence necessary for comprehension, and second, the temporal signature of the sequence, which falls approximately within the 4 Hz range (~250 ms duration of a typical syllable).

An experimental assessment of the effect of temporal distortion in the syllable structure to speech intelligibility has indicated that maximal comprehension occurred when the envelope of the compressed speech signal with included silence segments added up to ~4 Hz envelope overall (Ghitza & Greenberg, 2009). In the study, speech was compressed to three times normal speed. The compressed signal was then segmented into 40 ms intervals; periods of silence of varying duration (0–160 ms) were inserted between segments in two ways: periodically and aperiodically. Among all of the resulting stimuli of rapid speech segments with periodic versus aperiodic insertion of pauses of varying duration, intelligibility was highest for the stimulus with periodic insertion of 80 ms silences. The combination of a 40 ms speech segment and 80 ms silence (120 ms) constituted the Nyquist rate for a syllabic envelope, as two such segments with this length silences constitutes a 240 ms segment (close to the duration of the average syllable).

Other manipulations of the syllabic envelope of speech yield similar results. For example, flattening of the 4 Hz envelope of speech resulted in poor comprehension (Ghitza, 2012), while insertions of extra information, restricted to the input syllabic rhythm, improved intelligibility. Ghitza (2012) suggested that flattening the syllable envelopes prevented participants from tracking the input rhythm as syllabic, thus disrupting the hierarchical structure in the process of linguistic signal parsing. Any manipulation that restored the syllabic envelope in the input information resulted in the extraction of additional information from the modulation spectrum. This conclusion is corroborated by experimental manipulations in the spectral domain (Stilp & Kluender, 2010, 2016), which demonstrated that intelligibility dropped dramatically with the removal of changes in the spectral entropy in the signal, as opposed to removal of linguistically defined segments, such as vowels, consonants, or consonant-vowel transitions.

While behavioral evidence for intelligibility of variously manipulated speech stimuli is informative, a comprehensive picture of human communication is impossible without accounting for the neural mechanisms underlying comprehension in response to stimuli. The neural dimension of speech processing has been addressed primarily with the use of electrophysiology (EEG), which provides data with high temporal resolution (on the order of milliseconds). EEG signals can also be

analyzed in the spectral domain, like auditory speech, bringing the signal and neural response to it within the common quantitative framework.

Oscillatory activity of the brain, and the mechanisms of entrainment (to a stationary external stimulus) and phase re-setting (to an incoming stimulus) are crucial to both cognition and language (Giraud & Poeppel, 2012). Various frequencies of human brain activity, including delta (~2 Hz), theta (4 Hz), and gamma (<30 Hz), have been implicated in linguistic processing (Gross et al., 2013), which proceeds on multiple timescales simultaneously—compare, for example, the <1 Hz envelope of intonation versus ~4 Hz envelope of syllabic parsing (Blumenthal-Dramé & Malaia, 2018). However, multiple studies of intelligibility parameters indicate that the temporal granularity of 4 Hz is crucial to syllabic parsing, which appears to be an indispensable step in language comprehension.

The fluctuations of the syllable envelope appear to drive neural coupling in delta and theta bands (Doelling, Arnal, Ghitza, & Poeppel, 2014). Sharp changes ("edges") in the speech envelope have been shown to phase-reset auditory cortex oscillations, enhancing their entrainment to speech (Gross et al., 2013). Syllable rise-fall dynamics thus likely act as temporal clues that drive phase re-alignment in the brain activity, which allows parsing of the incoming signal into syllables. Analysis of the temporal modulation transfer function of human auditory perception (Edwards & Chang, 2013) identified a sensitivity peak in the syllabic range (~2–5 Hz), suggesting that envelope modulation measures for speech perception should also fall within that range. In the processing of the linguistic signal, further sampling and decoding, both in lower and higher frequency domains (e.g., prosody and phonological features), take place more efficiently when built upon the parsed syllables.

One theoretical viewpoint on the auditory processing of speech (Ghitza, 2011, 2013) suggests that emergence of theta-syllable[2] is driven by the organization of neurocomputational networks in the auditory cortex. Broadly, Ghitza (2013) suggests that speaking in timed packets (syllables) is the result of an evolutionary attempt to maximize information transfer (measured behaviorally by intelligibility) by way of matching the oscillatory rate underlying the cortical algorithm for auditory parsing. Amplitude modulations (speech envelope), Ghitza argues, enable reliable theta-driven parsing and decoding in the auditory cortex. The model, however, does not take into account the fact that the syllable duration in sign language, first, is equivalent to that of speech (Wilbur & Nolen, 1986), and second, elicits neural entrainment that is causally related to comprehension (Malaia, Borneman, Krebs, Roehm, & Wilbur, 2019). Abstracting from the input modality, the neurocomputational system for sign language does utilize brain regions that are typically associated with auditory processing, such as the temporal cortex. The right hemisphere, specifically right superior temporal gyrus (STG), has been implicated in sign syllable processing in functional neuroimaging research (Malaia et al., 2014; Malaia, Ranaweera, Wilbur, & Talavage, 2012). Voxelwise connectivity analyses (Malaia, Ikuta, & Wilbur, 2019) likewise indicate strong connections between right inferior frontal gyrus, and right STG in signers, which suggests that information flow between syllable-segmenting and the semantic system in sign language is at least partially right-lateralized. Thus, the proposal that the biological organization of human neurocomputational brain networks might contribute to the emergence of the theta-syllable is not without merit; however, the complete picture would have to account for modality-independent computations.

To see how comparable the sign signal is to speech, it is worth considering the way that syllable envelopes are manipulated in sign languages for semantic purposes. First, it should be mentioned that most lexical items (words) in sign languages studied to date are monosyllabic, parallel to spoken languages like Mandarin Chinese. A smaller portion are disyllabic, but with the caveat that the second syllable is rotated by 90 or 180° to the direction of movement of the first syllable. A third group consists of a single repetition of the first syllable (multiple repetitions are not lexical items but morphologically modified for purposes of verb aspect or [lack of] adjective intensity). Finally, there are compounds that are the result of reductions of two separate signs into the syllable requirements for single signs (Wilbur 2015). When signs are produced fluently in sentences, there are almost always transitional movements between them, for example, when one sign ends with the hand(s) located in one place and the next sign starts with them located somewhere else, there must be a movement of the hands to that next location before the next sign can start its lexical movement. These transitions are clearly differentiable to signers, and ignored when they are asked to count/tap to syllables (Wilbur & Allen, 1991; Wilbur & Nolen, 1986).

Given the large number of monosyllabic signs, it is instructive to observe movement modifications that can be made that affect the overall velocity envelope for different morpho-semantic purposes. In a typical single occurrence of, in essence, an unmarked verb (and ignoring all details related to handshape, place of articulation, number of hands, etc), the trajectory of the movement is relatively smooth, beginning with an acceleration of the movement, reaching a peak speed, and then decelerating smoothly; the trajectory is like a ball thrown into the air that goes up and eventually comes back down—a simple application of the physics of motion. Like the ballplayer who knows when the ball will come down, the signer knows about halfway through the sign how the sign will end. Verbs that have this type of movement trajectory envelope generally denote events that atelic, that is, have no natural end-state at which they could be considered finished/complete (Malaia, Borneman, & Wilbur,

2008; Malaia & Wilbur, 2012; Malaia, Wilbur, & Milkovic, 2013; Milkovic & Malaia, 2010; Wilbur & Malaia, 2008). In contrast, verbs that denote events that have distinct end-states (telics) are visually end-marked by a rapid deceleration to a stop, even when produced inside a sentence or narrative; this has been termed the Event Visibility Hypothesis Wilbur, 2008, 2010). These end-marked signs are the result of a change in muscle tension part-way through the hand movement trajectory, interrupting the smooth predictable rhythm. This end-marking (or lack thereof) is obvious to nonsigners (Strickland et al., 2015), providing evidence that the visual change detection algorithm is automatic. The same mechanism is used to make morphologically modified adjectives in (Wilbur, Malaia, & Shay, 2012) as well as being observed in other sign languages.

3.2 | Top-down processing: Syllable parsing as a gatekeeping mechanism for intelligibility

The relationship between syllabic parsing and top-down (attention and executive-function driven) processing is that of gatekeeping: multiple studies indicate that syllabic segmentation is an automatic perceptual process, which is independent of attention. However, disruptions in syllabic segmentation critically undermine both behaviorally measured comprehension, and neurally measured oscillatory activity (syllabic entrainment) in the theta band.

Interestingly, even disruptions to the spectral structure of the signal, which preserve the syllabic envelope fluctuations, do not affect intelligibility of the stimuli. Kayser, Ince, Gross, and Kayser (2015) demonstrated this phenomenon in interaction between top-down and bottom-up processes in speech perception by inserting pauses between syllables and words and disrupting speech rate, while preserving envelope fluctuations (entropy structure) of the signal. The manipulation did not affect the intelligibility of the stimuli or theta-band entrainment (4–8 Hz). Delta-band entrainment (2 Hz) was disrupted instead. This work demonstrated the functional distinction between syllabic, entropy-envelope-based entrainment in the theta band, and post-parsing processes, undergirded by oscillatory activity in lower and higher frequency bands. Ding et al. (2018) further demonstrated that while the process of converting the auditory signal into meaning relies on focused attention, parsing of the auditory signal happens automatically, whereas later speech-processing stages (e.g., lexical retrieval) are blocked by lack of attention.

Another type of change to the theta envelope of the speech signal, namely acceleration of syllabic rate, does result in adjustment of neural phase patterns to the accelerated syllabic rate (Pefkou, Arnal, Fontolan, & Giraud, 2017). With respect to the neural response to this manipulation, the comprehensibility of speech correlated with power in the low beta-band (14–21 Hz), suggesting that both bottom-up (syllable-driven parsing) and top-down (language-specific decoding of spectrotemporal data at higher frequencies) are required for intelligibility. Syllable parsing, in this case, also acted as a gatekeeping mechanism for other types of entrainment: it has to be accomplished first (in this case, via entrainment to the accelerated syllable rate), for the comprehension to occur.

Pitt, Szostak, and Dilley (2016) showed that only the speed (rhythm) of intelligible speech influenced whether a function word-syllable was heard in short word sequences (e.g., minor or child). The rate of degraded speech (low-pass filtered or sinewave), or tone sequences surrounding syllables in function words did not alter perception. Similar effects have been found for perception of weak syllables in both function words and lexical items (cease versus see us) (Baese-Berk, Dilley, Henry, Vinke, & Banzina, 2019), in native and nonnative speakers, as well as languages with distinctly different morphosyntactic properties, such as Russian and Mandarin (Dilley, Morrill, & Banzina, 2013; Lai & Dilley, 2016).

A similar phenomenon resulting from long-term effect of language exposure is observable is signers' versus nonsigners perception of sign language. Like spoken syllables, signed syllables can be counted and tapped to, and native deaf signers, hearing native signers, and hearing subjects with no sign familiarity perform differently on this task (Allen et al., 1991). Although the three groups showed comparable rhythmic tapping to repeated signs, and to signs with primary stress, the two fluent signing groups tapped less to signs with secondary stress and unstressed signs than did the sign naïve group. This data demonstrate that knowledge of sign language alters appreciation of the rhythmic structure of visual input—a phenomenon similar to the top-down effects observed in speech.

While no equivalent research on neural dependency of top-down and bottom-up processing on syllabic entrainment is yet available for sign language, a patchwork of studies appears to indicate that a similar segmentation-to-comprehension processing sequence exists in the visual linguistic modality. For example, Brookshire, Lu, Nusbaum, Goldin-Meadow, and Casasanto (2017) has shown a cortical entrainment to visual stimuli in sign language at <5 Hz.[3] Coherence to sign stimuli was reported to be strongest over occipital and parietal cortex, as expected for those who use the visual modality for communication (Malaia & Wilbur, 2010, 2018). Hearing nonsigner participants in the Brookshire et al. (2017) study also demonstrated coherence to visual sign language stimuli, but for them, the entrainment at frontal sites was reduced relative to the fluent signers, which is an expected finding, given the lack of comprehension in this group.

Approaching the question of syllable salience from the production angle, an analysis of interference effects in Catalan Sign Language (Baus, Gutiérrez, & Carreiras, 2014) has shown that distractor signs that shared the syllable-forming Movement feature with target signs affected signing latencies more than other shared feature combinations. In this study, a picture-sign interference task, native and nonnative signers were asked to sign picture names while ignoring distractor-signs, with which the target signs shared two phonological parameters out of three (*Location, Movement*, and *Handshape*). Movement appeared as both the strongest potential distractor, when handshape feature was also shared (although only for nonnative signers), and as the strongest facilitator, where pictures and distractors shared Location-Movement. The authors concluded that the results indicated perceptual salience of the syllabic structure to signers, although it was not possible to make a distinction between whether this salience was due to modality, or to native language effect, on the basis of the data (for discussion of visual-graphemic and sign parameter interference in signer memory, see Bernstein Ratner & Wilbur, 1984; and further signer memory discussion in Malaia & Wilbur, 2018).

Corroborating evidence for the salience of syllabic frequency (4 Hz) in both visual and auditory modalities comes from speech research investigating cross-modal integration in audiovisual speech. Congruent audiovisual speech enhances our ability to comprehend a speaker, especially in noisy conditions. Crosse, Butler, and Lalor (2015) EEG analysis of audiovisual integration found that the effect was most prominent at the temporal scale corresponding to syllabic rate (2–6 Hz) (Crosse et al., 2015). In later work, the same group (Crosse, Di Liberto, & Lalor, 2016) found that the improvement in neural tracking of the audio speech signal in noisy listening conditions by addition of congruent visual input predicted the multisensory gain in behavioral performance at a time lag of 250 ms (4 Hz). At the same time, however, crossmodal integration over long temporal windows also contributed to increased comprehension. These findings suggested that integration mechanisms that contribute to the efficient processing of audiovisual speech in background noise include both those that operate at the syllabic rate (4 Hz), and those that operate at lower frequencies, taking visual cues for higher-level linguistic structures (phrases, sentences) into account.

3.3 | Bottom-up effects of spectro-temporal context in which syllable occurs

So far, we have considered cross-modal and cross-linguistic universals in syllabic parsing, which consist of the following:

1. The necessity for the rise-fall pattern in the signal entropy (auditory or visual), and its appropriate temporal granularity (~4 Hz);
2. Primacy of automatic parsing of the input signal based on entropy rise-fall pattern, which triggers neural entrainment even without focused attention;
3. The critical role of theta-entrainment (syllabic parsing) in development of neural entrainment at other scales, and comprehension of the signal.

Beyond these universals, however, there also exists substantial variability in how easily users of specific languages are primed by the temporal patterns in entropy variations. Generally speaking, individual linguistic experience determines the primary strategy of syllable parsing, and subsequent comprehension. For example, consider these two sentences: *Jill got quite mad when she heard* **there are** *birds* and *Jill got quite mad when she heard* **their** *birds*. Both are grammatical; moreover, the pronunciation can be nearly identical. When listeners hear sentences that have a grammatical interpretation with or without a critical function word and transcribe what they heard (Morrill, Dilley, McAuley, & Pitt, 2014), they are more likely to report critical function words (*there are*) when the repeating syllabic pattern in the text preceding the critical words matched the rhythm in the function-word-containing region. Such effects of the primed syllabic parsing on the amount of lexical material listeners extract highlights the importance of syllable timing for parsing strategy even within the native language (see also Box 1 for the relevance of syllable timing in neurally based language disorders). The phenomenon of syllabic entrainment causally affecting information decoding for incoming speech has also been observed in a vowel duration study (Kösem et al., 2018), where syllabic entrainment persisted for several cycles after speech rhythm changes. Sustained entrainment to faster rhythm resulted in perceiving the duration of a vowel as longer, thereby changing word meaning.

Heffner, Dilley, McAuley, and Pitt (2013) demonstrated an effect of surrounding spectro-temporal context (speech rate) on syllable onset perception. Speech rate factor interacted with local acoustic parameters indicative of syllable onset: intensity (Experiment 1), fundamental frequency (Experiment 2), word duration (Experiment 3), and high frequency noise resembling a consonantal onset (Experiment 4) such that both local and contextual cues had independent gradient effects, indicative of probabilistic, rather than absolute approaches to syllabification.

BOX 1   **Disruptions in syllable entrainment result in atypical language processing**

Investigations of atypical language processing in neurodevelopmental and neurodegenerative disorders frequently identify difficulties in temporal entrainment to syllable frequency as causal for comprehension difficulties. Leong and Goswami (2014) tested both perception and production of syllabic rhythm in adults with dyslexia. In perception task, adults tapped rhythmically to nursery rhyme sentences. In the perception task, the participants produced the same rhymes out loud to the metronome. Rhythmic entrainment, assessed via rhythmic indices at various timescales, revealed irregularities in the frequency range corresponding to syllabic patterning (~4–5 Hz) in both perception and production. The authors concluded that syllable timing deficits disrupt phonological representations for spoken words during development, suggesting that this disruption constitutes the central cognitive characteristic of developmental dyslexia cross-linguistically.

In children with dyslexia, oscillatory entrainment to auditory, visual, and audio-visual speech is also impaired. In an EEG study, children with dyslexia differed in neural phase alignment with the incoming signal; the differences in alignment were also correlated with behavioral measures of reading (Power, Mead, Barnes, & Goswami, 2013). The authors proposed that disruption in phase encoding to speech at low frequencies underlies cognitive impairment in dyslexia. More generally, reading ability in pre-schoolers appears to strongly correlate with the ability to neurally synchronize to an external beat, which, in turn, relates to ability to encode speech. The pre-schoolers able to synchronize their neural activity to an external beat performed better on reading-related tasks measuring phonological processing, auditory short-term memory, and rapid naming (Woodruff Carr, White-Schwoch, Tierney, Strait, & Kraus, 2014). Thus, neural encoding of temporal modulations in auditory stimulus appeared to be one of the key mechanisms underlying reading acquisition.

In Parkinson's disease the ability to initiate speech (measured via turn-taking delay) and alignment of speech rhythm with stimuli is also disrupted. However, individuals with Parkinson's appear to perform better when metric rhythm is available from the stimuli, exploiting regular cues in speech (Späth et al., 2016). This underscores the role of the syllable-level envelope in temporal organization of both speech production and perception mechanisms.

Notably, the listeners appear to track the pace of speech of their interlocutors over an extended duration and it is this global speech rate that affects which words listeners hear (Baese-Berk et al., 2014). A real-life equivalent of this might be someone preparing to go to New York, and mentally adjusting to rapid-fire interlocution based on prior experiences with the population of the city.

The patterns of syllabic entrainment interact subtly in cross-modal phenomena, such as the McGurk effect.[4] Investigations of ambiguous stimuli /ba/ and /ga/ in behavioral and fMRI paradigms (Ten Oever et al., 2016; ten Oever & Sack, 2015) have shown that correlation of the stimulus phase with the phase of underlying neural activity determines behavioral distinction between the syllables. Specifically, the 80-ms delay[5] in voice onset time (VOT[6]) for /ga/, as compared with /ba/ appears to be identified by way of phase synchronization. The measured phase difference in which perception is biased toward /da/ or /ga/ exactly matched the difference in the temporal onset delays in natural audiovisual speech between mouth movements and speech sounds, which last 80 ms longer for /ga/ than for /da/. These results indicate the functional relationship between pre-stimulus phase and syllable identification, and signify that the origin of this phase relationship could lie in exposure to and subsequent learning of unique audiovisual temporal onset differences of different syllables. At the neural level, a phase lag manipulation of neural response to speech in the STG using transcranial alternating current stimulation affected the ability to detect temporal irregularities in intelligible speech (Zoefel, Archer-Boyd, & Davis, 2018), suggesting that STG plays a critical role in neurocomputational analysis of temporal parameters of the speech signal.

In the domain of sign languages (fully visual communication), the variability of motion in sign-syllables forms the basis of the quantitative distinction between noninformative, biological motion, and the sign language signal (Borneman, Malaia, & Wilbur, 2018; Malaia et al., 2016). In general terms, the mathematically quantified amount of information (i.e., variability) in the motion of the articulators in sign language forms the basis of sign syllables. Different experimental approaches, including video analysis using optical flow metric and motion capture data analysis, indicate that information transfer in sign language critically relies on the entropy of the visual signal (Borneman et al., 2018; Malaia, Borneman, & Wilbur, 2017). In general terms, the temporal structure of increases and decreases in hand-and-body motion velocity forms the basis of the visual envelope used to track syllabic structure in sign language (see Figure 1).
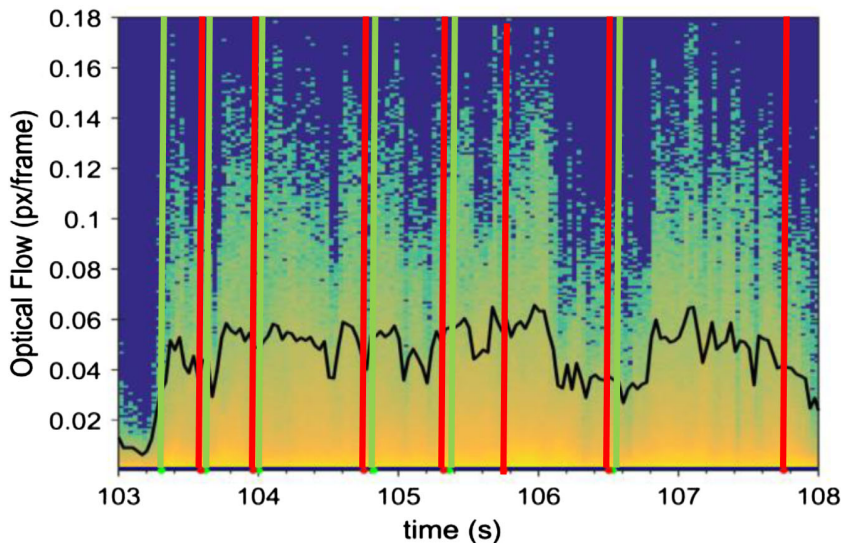
FIGURE 1 The overall optical flow across velocity bins (plotted on the *y*-axis) and time (*x*-axis) in a two-dimensional video of a signed sentence in ASL. The integrated sum of velocities across bins is plotted as a black line. Green vertical lines mark word onsets; red vertical lines mark word offsets, which are followed by transitional motion

IF MEET ALL THEN PROGRESS TAKE-UP EVALUATION
If (you) meet all (criteria for participation) then we move on to evaluation.

A substantial body of research indicates that signers employ entropy variations in the visual signal both in production and perception (Klima, Tzeng, Fok, Bellugi, & Corina, 1996; Malaia et al., 2016). A neuroimaging investigation of comprehension of ASL signs differing in syllable structure in Deaf participants identified that processing differences stem from the neural regions involved in syllable timing—the right STG-cerebellar network.[7] Further, comparison of neural activations of signers and nonsigners who were shown these ASL signs indicated clear differences between linguistic processing of the syllabic structure by signers, and processing of the same visual signal without linguistic comprehension by nonsigners (Malaia et al., 2012). Signers, who understood ASL, processed syllable-structure differences between verb signs, and extracted semantic information for comprehension. For nonsigners, however, the two types of verb signs only appeared to have velocity-based differences, as indicated by differences in the occipito-temporal junction (area MT+). No integration with any other level of linguistic processing, or comprehension, was possible for nonsigners who were unfamiliar with American Sign Language. The differences between processing of the same signal (ASL) by signers and nonsigners lie in the fact that signers could combine the perceived change-units (syllables) into hierarchical structures for syntactic processing, and relate them to items stored in memory (lexical processing), whereas in contrast, nonsigners stopped at perceptual segmentation of visual events (Malaia & Wilbur, 2010; Malaia, Wilbur, & Weber-Fox, 2013).

The mechanism of neural entrainment to entropy differentials that users of spoken and signed languages employ for language comprehension appears to be similar, regardless of communication modality. In EEG data of signers viewing signed sentences (meaningful stimuli), and the same sentences played in reverse (meaningless stimuli with rich spectrotemporal structure), the maximum coherence between the entropy variations in the optical flow of the visual stimuli and neural activity of the participants occurred between 100 ms and 250 ms post-stimulus onset (Malaia et al., under review). In response to meaningful signed sentences, coherence values were higher over fronto-central electrode sites; in response to time-reversed (and, therefore, meaningless) videos, negative values of cross-correlation were observed. The results indicate that comprehension of signed sentences is based on entrainment of neural oscillations to the dynamic variations in spatial entropy of the visual signal. The findings demonstrate that the cortical tracking of spectro-temporal entropy of the signal is a modality-independent mechanism for event segmentation for communication, or syllable parsing, in humans. Such results point to the likelihood of modality-independent evolution of language based on cortical entrainment that facilitates scene segmentation, and action perception and production (Blumenthal-Dramé & Malaia, 2018).

We propose that the perceptual entropy-tracking mechanism operating at ~4 Hz is modality-independent, and has, over time, led to universal production of human linguistic communication in time-locking patterns at the same frequency, conventionally understood as syllable. Figure 2 illustrates the primary role of syllabic parsing in language comprehension, as reviewed in this manuscript.

While this review primarily focused on the similarities in bottom-up (signal-driven) effects in visual and auditory modalities in language processing, the top-down effects that are modulated by longer-term exposure (either language knowledge, or speech rate over longer periods of time) clearly contribute to "tuning" of both spectro-temporal perceptual filters. Top-down

FIGURE 2 The Entropy Syllable Parsing model as a modality-independent signal segmentation framework for communication: The envelope of the incoming signal in any modality is used as the basis for its perceptual parsing. Parsing at lower and higher frequencies is subsequent to syllabic parsing
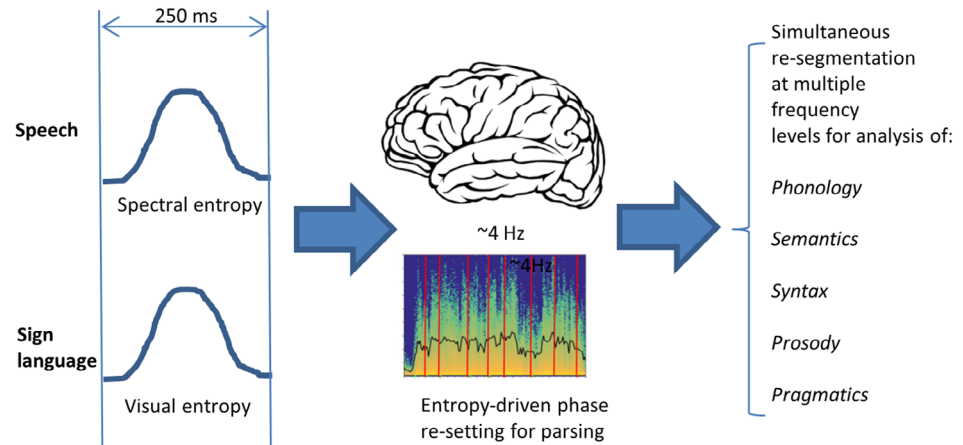
effects are captured by theoretic frameworks that emphasize the role of prediction in the brain (predictive processing, Clark, 2013; probabilistic/Bayesian inference, Griffiths, Chater, Kemp, Perfors, & Tenenbaum, 2010; free energy principle, Friston & Kiebel, 2009, Friston, 2010). Predictive processing models draw on experimental results indicating that different sensory regions of the brain encode not the external sensory signal, but instead, the "prediction error": the difference between a predicted external signal, and the actual external signal (Lee & Mumford, 2003; Norris, McQueen, & Cutler, 2016; Spratling, 2008a, 2008b, 2016). Multiple algorithms implementing predictive processing have been proposed (see Spratling, 2017 for review of similarities and differences among predictive coding implementations). In contrast, probabilistic (Bayesian) inference frameworks aim to represent not a processing algorithm, but a more general goal-oriented behavior, the result of which is a mental state that represents latent causes of the observed sensory input (see Aitchison & Lengyel, 2017, for comparison between predictive processing and Bayesian inference). Empirical research in cross-modal universals of language comprehension will still need to resolve quantitative parameters of between-scale effects, and possibly quantify them for specific populations based on the neural malleability thresholds. For example, one could pose a question such as "how much exposure to sign language can alter temporal parameters of visual scene perception in 20-year-old nonsigners, and for how long will the effect persist?" Multiple data points of this type, combined with understanding of individual variability of response to exposure to auditory and visual stimuli can help, over time, create a framework that might help us understand how humans respond to long-term changes in their visual and auditory environment.

4 | CONCLUSION

Throughout the history of linguistics, approaches to syllable definition have been proposed based on articulation, acoustics, and phonological characteristics of the signal, such as the structuring of sequences of segments, and the location of boundaries between syllables. The various contributors to these debates have argued, or assumed, that even if the syllable is not a necessary concept, it is certainly a preferred one. We advance a framework of syllable definition based on information theory application to communication: Entropy Syllable Parsing model. It goes beyond a single modality, and brings together convergent evidence from neuroscience of language perception and production, by showing that entropy changes in the linguistic signal underlie perceptual syllable segmentation in both speech and sign.

Ability to parse syllables—dynamic changes in signal entropy—from connected input is the modality-independent parameter that is crucial to language proficiency. The Entropy Syllable Parsing model predicts segmentation of the linguistic signal in both auditory and visual modalities as reliant on algorithmically similar neural underpinnings, and manifesting in similar behavioral performance.

CONFLICT OF INTEREST

The authors have declared no conflicts of interest for this article.

ORCID

Evie A. Malaia https://orcid.org/0000-0002-4700-0257

ENDNOTES

[1] A full phonological description is provided by Brentari (1998).

[2] Referring to theta rate, or ~4 Hz, at which the most typical 250 ms syllable takes place.

[3] The dynamic changes in the visual stimuli in this study were quantified based on the amount of color changes between frames, rather than a measure of articulator movement. While the cortical location of entrainment observed in the study is of interest, the frequency band of reported coherence is not linguistically meaningful.

[4] The McGurk effect is a cross-modal illusion, whereby the visual information one receives when seeing a person speak (in addition to auditory information) changes what sound is perceived to have been pronounced (McGurk & MacDonald, 1976; also Miller, 1981; Theodore, Miller, & DeSteno, 2009). The effect is based on audiovisual integration ability which is critical for normal language processing; the McGurk effect is significantly reduced in populations with disorders of speech and language (dyslexia, SLI, autism spectrum disorder).

[5] The amount of delay is a specific value within the study; in general, as mentioned above, it depends on the context speech rate.

[6] Voice onset time is the length of time that passes between the release of a stop consonant and the onset of voicing.

[7] Since a significant amount of linguistic processing for sign languages occurs in the right hemisphere, right hemispheric lateralization for this difference is backed by multiple independent experiments (see Malaia et al., 2014, for summary).

ORCID

Evie A. Malaia https://orcid.org/0000-0002-4700-0257

RELATED WIREs ARTICLES

One grammar or two? Sign Languages and the Nature of Human Language
Understanding the developing sound system: interactions between sounds and words.
Visual sign phonology: Insights into human reading and language from a natural soundless phonology.
Predictive coding
Shared neural and cognitive mechanisms in action and language: The multiscale information transfer framework

FURTHER READING

Malaia, E. (2014). It still isn't over: Event boundaries in language and perception. *Language and Linguistics Compass*, 8(3), 89–98.
van der Hulst, H. (1995). Head-dependency relations in the representation of signs. In H. Bos & T. Schermer (Eds.), *Sign language research 1994: Proceedings of the fourth European congress on sign language research* (pp. 11–38). Hamburg, Germany: Signum Press.

REFERENCES

Aitchison, L., & Lengyel, M. (2017). With or without you: Predictive coding and Bayesian inference in the brain. *Current Opinion in Neurobiology*, 46, 219–227.
Allen, G. D., Wilbur, R. B., & Schick, B. B. (1991). Aspects of rhythm in ASL. *Sign Language Studies*, 72(1), 297–320.
Anderson, S. R. (1974). *The organization of phonology*. New York, NY: Academic Press.
Baese-Berk, M. M., Dilley, L. C., Henry, M. J., Vinke, L., & Banzina, E. (2019). Not just a function of function words: Distal speech rate influences perception of prosodically weak syllables. *Attention, Perception, & Psychophysics*, 81(2), 571–589.
Baese-Berk, M. M., Heffner, C. C., Dilley, L. C., Pitt, M. A., Morrill, T. H., & McAuley, J. D. (2014). Long-term temporal tracking of speech rate affects spoken-word recognition. *Psychological Science*, 25(8), 1546–1553.
Baldwin, D. A., Baird, J. A., Saylor, M. M., & Clark, M. A. (2001). Infants parse dynamic action. *Child development*, 72(3), 708–717.
Battison, R. (1978). *Lexical borrowing in American sign language*. Silver Spring, MD: Linstok.

Baus, C., Gutiérrez, E., & Carreiras, M. (2014). The role of syllables in sign language production. *Frontiers in Psychology*, *5*, 1254.

Bellugi, U., & Klima, E. S. (1979). The signs of language. *Psychological Science*, *1*, 6–9.

Bernstein Ratner, N., & Wilbur, R. B. (1984). Another look at memory strategies in deaf signers. *Journal of Visual Verbal Languaging*, *4*(2), 55–63.

Blumenthal-Dramé, A., & Malaia, E. (2018). Shared neural and cognitive mechanisms in action and language: The multiscale information transfer framework. *WIREs Cognitive Science*, *10*, e1484. https://doi.org/10.1002/wcs.1484

Borneman, J. D., Malaia, E. A., & Wilbur, R. B. (2018). Motion characterization using optical flow and fractal complexity. *Journal of Electronic Imaging*, *27*(05), 1. https://doi.org/10.1117/1.JEI.27.5.051229

Brentari, D. (1998). *A prosodic model of sign language phonology*. Cambridge, MA: MIT Press.

Brookshire, G., Lu, J., Nusbaum, H. C., Goldin-Meadow, S., & Casasanto, D. (2017). Visual cortex entrains to sign language. *Proceedings of the National Academy of Sciences*, *114*(24), 6352–6357.

Cairns, C. E., & Feinstein, M. H. (1982). Markedness and the theory of syllable structure. *Linguistic Inquiry*, *13*(2), 193–225.

Chinchor, N. (1981). *Numeral incorporation in American Sign language*. (Unpublished doctoral dissertation). Brown University, Providence.

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

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(03), 181–204. https://doi.org/10.1017/S0140525X12000477

Clements, G. N., & Keyser, S. J. (1983). CV phonology. A generative theory of the syllable. *Linguistic Inquiry Monographs 9.* Cambridge, MA: MIT Press.

Corina, D. P. (1990). *Handshape assimilations in hierarchical phonological representation*. Sign Language Research: Theoretical Issues, 27–49.

Crosse, M. J., Butler, J. S., & Lalor, E. C. (2015). Congruent visual speech enhances cortical entrainment to continuous auditory speech in noise-free conditions. *Journal of Neuroscience*, *35*(42), 14195–14204.

Crosse, M. J., Di Liberto, G. M., & Lalor, E. C. (2016). Eye can hear clearly now: inverse effectiveness in natural audio visual speech processing relies on long-term cross modal temporal integration. *Journal of Neuroscience*, *36*(38), 9888–9895.

Dilley, L., Morrill, T., & Banzina, E. (2013). New tests of the distal speech rate effect: Examining cross-linguistic generalization. *Frontiers in Psychology*, *4*, 1002.

Ding, N., Pan, X., Luo, C., Su, N., Zhang, W., & Zhang, J. (2018). Attention is required for knowledge-based sequential grouping: Insights from the integration of syllables into words. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *38*(5), 1178–1188. https://doi.org/10.1523/JNEUROSCI.2606-17.2017

Doelling, K. B., Arnal, L. H., Ghitza, O., & Poeppel, D. (2014). Acoustic landmarks drive delta–theta oscillations to enable speech comprehension by facilitating perceptual parsing. *NeuroImage*, *85*, 761–768. https://doi.org/10.1016/j.neuroimage.2013.06.035

Edwards, E., & Chang, E. F. (2013). Syllabic (∼2-5 Hz) and fluctuation (∼1-10 Hz) ranges in speech and auditory processing. *Hearing Research*, *305*, 113–134. https://doi.org/10.1016/j.heares.2013.08.017

Friedman, L. (1974). *On the physical manifestation of stress in the American sign language*. (Unpublished manuscript). University of California, Berkeley.

Friedman, L. (1976). Phonology of a soundless language: Phonological structure of the American Sign Language. Berkeley, CA: UC Berkeley dissertation.

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, *11*(2), 127–138.

Friston, K., & Kiebel, S. (2009). Predictive coding under the free-energy principle. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1521), 1211–1221.

Ghitza, O. (2011). Linking speech perception and neurophysiology: Speech decoding guided by cascaded oscillators locked to the input rhythm. *Frontiers in Psychology*, *2*. https://doi.org/10.3389/fpsyg.2011.00130

Ghitza, O. (2012). On the role of theta-driven syllabic parsing in decoding speech: Intelligibility of speech with a manipulated modulation spectrum. *Frontiers in Psychology*, *3*, 238. https://doi.org/10.3389/fpsyg.2012.00238

Ghitza, O. (2013). The theta-syllable: A unit of speech information defined by cortical function. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00138

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*(1–2), 113–126. https://doi.org/10.1159/000208934

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

Goldsmith, J. (1979). *Autosegmental and metrical phonology*. Oxford, England: Blackwell.

Goldstein, L., & Browman, C. P. (1986). Representation of voicing contrasts using articulatory gestures. *Journal of Phonetics*, *14*(2), 339–342.

Griffiths, T. L., Chater, N., Kemp, C., Perfors, A., & Tenenbaum, J. B. (2010). Probabilistic models of cognition: Exploring representations and inductive biases. *Trends in Cognitive Sciences*, *14*(8), 357–364.

Gross, J., Hoogenboom, N., Thut, G., Schyns, P., Panzeri, S., Belin, P., & Garrod, S. (2013). Speech rhythms and multiplexed oscillatory sensory coding in the human brain. *PLoS Biology*, *11*(12), e1001752. https://doi.org/10.1371/journal.pbio.1001752

Heffner, C. C., Dilley, L. C., McAuley, J. D., & Pitt, M. A. (2013). When cues combine: How distal and proximal acoustic cues are integrated in word segmentation. *Language and Cognitive Processes*, *28*(9), 1275–1302.

Hooper, J. (1972). The syllable in generative phonology. *Language*, *48*, 525–540.

Hyman, L. M. (1975). *Phonology: Theory and analysis*. New York, NY: Harcourt College Pub.

Hyman, L. M. (1985). *A theory of phonological weight*. Dordrecht: Foris.

Kahn, D. (1976). *Syllable-based generalizations in English phonology*. Cambridge, MA: MIT Press.

Kayser, S. J., Ince, R. A. A., Gross, J., & Kayser, C. (2015). Irregular speech rate dissociates auditory cortical entrainment, evoked responses, and frontal alpha. *Journal of Neuroscience*, *35*(44), 14691–14701. https://doi.org/10.1523/JNEUROSCI.2243-15.2015

Kegl, J. A., & Wilbur, R. B. (1976). When Does Structure Stop and Style Begin? Syntax, Morphology, and Phonology v.-s. Stylistic Variation in American Sign Language. Papers from the Regional Meeting. Chicago Linguistic Society, *12*, 376–396.

Kenstowicz, M. & Kisseberth, C. (1979). *Generative phonology: Description and theory*. New York: Academic Press.

Klima, E. S., Tzeng, O., Fok, A., Bellugi, U., & Corina, D. (1996). From sign to script: Effects of linguistic experience on perceptual categorization (Tech. Rep. No. INC-9604). San Diego: University of California, San Diego, Institute for Neural Computation.

Kluender, K. R., & Kiefte, M. (2006). Speech perception within a biologically realistic information-theoretic framework. In M. A. Gernsbacher, & M. Traxler (Eds.), *Handbook of psycholinguistics* (2nd ed., pp. 153–199). London, England: Elsevier.

Kluender, K. R., Stilp, C. E., & Lucas, F. L. (2019). Long-standing problems in speech perception dissolve within an information-theoretic perspective. *Attention, Perception, & Psychophysics*, *81*(4), 861–883.

Kösem, A., Bosker, H. R., Takashima, A., Meyer, A., Jensen, O., & Hagoort, P. (2018). Neural entrainment determines the words we hear. *Current Biology*, *28*(18), 2867–2875.e3. https://doi.org/10.1016/j.cub.2018.07.023

Ladefoged, P. (1982). *A course in phonetics*. New York: Harcourt Brace Jovanovich.

Lai, W., & Dilley, L. (2016). Cross-linguistic generalization of the distal rate effect: Speech rate in context affects whether listeners hear a function word in Chinese Mandarin. *Proceedings of Speech Prosody*, *8*, 1124–1128.

Lass, R. (1984). *Phonology: An introduction to basic concepts*. Cambridge, MA: Cambridge University Press.

Lee, T. S., & Mumford, D. (2003). Hierarchical Bayesian inference in the visual cortex. *JOSAA*, *20*(7), 1434–1448.

Leong, V., & Goswami, U. (2014). Assessment of rhythmic entrainment at multiple timescales in dyslexia: Evidence for disruption to syllable timing. *Hearing Research*, *308*, 141–161. https://doi.org/10.1016/j.heares.2013.07.015

Lewicki, M. S. (2010). Information theory: A signal take on speech. *Nature*, *466*(7308), 821–822.

Liddell, S. K. (1984). THINK and BELIEVE: sequentiality in American Sign Language. *Language*, *60*(2), 372–399.

Liddell, S. K., & Johnson, R. E. (1989). American sign language: The phonological base. *Sign Language Studies*, *64*(1), 195–277.

Lidji, P., Palmer, C., Peretz, I., & Morningstar, M. (2011). Listeners feel the beat: Entrainment to English and French speech rhythms. *Psychonomic Bulletin & Review*, *18*(6), 1035–1041. https://doi.org/10.3758/s13423-011-0163-0

Loucks, J., & Baldwin, D. (2009). Sources of information for discriminating dynamic human actions. *Cognition*, *111*(1), 84–97.

Malaia, E., Borneman, J., & Wilbur, R. B. (2008). *Analysis of ASL motion capture data towards identification of verb type*. Proceedings of the 2008 Conference on Semantics in Text Processing, 155–164. Association for Computational Linguistics.

Malaia, E., Borneman, J. D., & Wilbur, R. B. (2016). Assessment of information content in visual signal: Analysis of optical flow fractal complexity. *Visual Cognition*, *24*(3), 246–251.

Malaia, E., Borneman, J. D., & Wilbur, R. B. (2017). Information transfer capacity of articulators in American sign language. *Language and Speech*, *61*(1), 97–112. https://doi.org/10.1177/0023830917708461

Malaia, E., Borneman, J., Krebs, J., Roehm, D., Wilbur, R. (2019). Cortical entrainment to visual entropy underlies sign language comprehension. *Poster presented at Theoretical Issues in Sign Language Research 13 (TISLR13)*, Hamburg, Germany.

Malaia, E., Ikuta, T., Wilbur, R. (2019). Neurofunctional architecture underlying syntactic processing in American Sign Language. *Poster presented at International Brain and Syntax Think Tank*, Northwestern University, IL.

Malaia, E., Ranaweera, R., Wilbur, R. B., & Talavage, T. M. (2012). Event segmentation in a visual language: Neural bases of processing American sign language predicates. *NeuroImage*, *59*(4), 4094–4101. https://doi.org/10.1016/j.neuroimage.2011.10.034

Malaia, E., Talavage, T. M., & Wilbur, R. B. (2014). Functional connectivity in task-negative network of the deaf: Effects of sign language experience. *PeerJ*, *2*, e446. https://doi.org/10.7717/peerj.446

Malaia, E., & Wilbur, R. B. (2010). Early acquisition of sign language: What neuroimaging data tell us. *Sign Language and Linguistics*, *13*(2), 183–199.

Malaia, E., & Wilbur, R. B. (2012). Kinematic signatures of telic and atelic events in ASL predicates. *Language and Speech*, *55*(Pt 3), 407–421. https://doi.org/10.1177/0023830911422201

Malaia, E., & Wilbur, R. B. (2018). Visual and linguistic components of short-term memory: Generalized neural model (GNM) for spoken and sign languages. *Cortex*, *112*, 69–79. https://doi.org/10.1016/j.cortex.2018.05.020

Malaia, E., Wilbur, R. B., & Milkovic, M. (2013). Kinematic parameters of signed verbs. *Journal of Speech, Language, and Hearing Research*, *56*(5), 1677–1688. https://doi.org/10.1044/1092-4388(2013/12-0257)

Malaia, E., Wilbur, R. B., & Weber-Fox, C. (2009). ERP evidence for telicity effects on syntactic processing in garden-path sentences. *Brain and Language*, *108*(3), 145–158. https://doi.org/10.1016/j.bandl.2008.09.003

Malaia, E., Wilbur, R. B., & Weber-Fox, C. (2013). Event end-point primes the Undergoer argument: Neurobiological bases of event structure processing. In B. Arsenijević, B. Gehrke, & R. Marín (Eds.), *Studies in the composition and decomposition of event predicates* (pp. 231–248). Dordrecht: Springer.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*(5588), 746–748.

Milkovic, M., & Malaia, E. (2010). *Event visibility in Croatian sign language: Separating aspect and aktionsart*. Poster session presented at the Tenth Meeting of Theoretical Issues in Sign Language Research, 165–166.

Miller, J. L. (1981). Effects of speaking rate on segmental distinctions. In Eimas & Miller (Eds.), *Perspectives on the Study of Speech*, 39–74.

Morrill, T. H., Dilley, L. C., McAuley, J. D., & Pitt, M. A. (2014). Distal rhythm influences whether or not listeners hear a word in continuous speech: Support for a perceptual grouping hypothesis. *Cognition*, *131*(1), 69–74. https://doi.org/10.1016/j.cognition.2013.12.006

Newkirk, D. (1979). *The form of the continuative aspect on ASL verbs*. (Unpublished Ms). The Salk Institute for Biological Studies, La Jolla, CA.

Newkirk, D. E. (1998). On the temporal segmentation of movement in American sign language. *Sign Language & Linguistics*, *1*(2), 173–211.

Norris, D., McQueen, J. M., & Cutler, A. (2016). Prediction, Bayesian inference and feedback in speech recognition. *Language, Cognition and Neuroscience*, *31*(1), 4–18. https://doi.org/10.1080/23273798.2015.1081703

Olofson, E. L., & Baldwin, D. (2011). Infants recognize similar goals across dissimilar actions involving object manipulation. *Cognition*, *118*(2), 258–264.

Oxenham, A. J., Boucher, J. E., & Kreft, H. A. (2017). Speech intelligibility is best predicted by intensity, not cochlea-scaled entropy. *The Journal of the Acoustical Society of America*, *142*(3), EL264–EL269. https://doi.org/10.1121/1.5002149

Padden, C. A., & Perlmutter, D. M. (1987). American sign language and the architecture of phonological theory. *Natural Language & Linguistic Theory*, *5*(3), 335–375.

Pefkou, M., Arnal, L. H., Fontolan, L., & Giraud, A.-L. (2017). θ-Band and β-band neural activity reflects independent syllable tracking and comprehension of time-compressed speech. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *37*(33), 7930–7938. https://doi.org/10.1523/JNEUROSCI.2882-16.2017

Pitt, M. A., Szostak, C., & Dilley, L. C. (2016). Rate dependent speech processing can be speech specific: Evidence from the perceptual disappearance of words under changes in context speech rate. *Attention, Perception, & Psychophysics*, *78*(1), 334–345.

Power, A. J., Mead, N., Barnes, L., & Goswami, U. (2013). Neural entrainment to rhythmic speech in children with developmental dyslexia. *Frontiers in Human Neuroscience*, *7*. https://doi.org/10.3389/fnhum.2013.00777

Sandler, W. (1986). The spreading hand autosegment of American sign language. *Sign Language Studies*, *1050*, 1–28.

Sandler, W. (1989). *Phonological representation of the sign: Linearity and nonlinearity in sign language phonology*. Dordrecht, The Netherlands: Foris.

Sargent, J. Q., Zacks, J. M., Hambrick, D. Z., Zacks, R. T., Kurby, C. A., Bailey, H. R., … Beck, T. M. (2013). Event segmentation ability uniquely predicts event memory. *Cognition*, *129*(2), 241–255.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, *27*(3), 379–423.

Shu, Y., Feng, X., & Chen, F. (2016). Comparing the perceptual contributions of cochlear-scaled entropy and speech level. *The Journal of the Acoustical Society of America*, *140*(6), EL517–EL521. https://doi.org/10.1121/1.4971879

Siple, P. (1978). Visual constraints for sign language communication. *Sign Language Studies*, *19*(1), 95–110.

Späth, M., Aichert, I., Ceballos-Baumann, A. O., Wagner-Sonntag, E., Miller, N., & Ziegler, W. (2016). Entraining with another person's speech rhythm: Evidence from healthy speakers and individuals with Parkinson's disease. *Clinical Linguistics & Phonetics*, *30*(1), 68–85. https://doi.org/10.3109/02699206.2015.1115129

Speer, N. K., Zacks, J. M., & Reynolds, J. R. (2007). Human brain activity time-locked to narrative event boundaries. *Psychological Science*, *18*(5), 449–455.

Spratling, M. W. (2008a). Predictive coding as a model of biased competition in visual attention. *Vision Research*, *48*(12), 1391–1408.

Spratling, M. W. (2008b). Reconciling predictive coding and biased competition models of cortical function. *Frontiers in Computational Neuroscience*, *2*, 4.

Spratling, M. W. (2016). Predictive coding as a model of cognition. *Cognitive Processing*, *17*(3), 279–305.

Spratling, M. W. (2017). A review of predictive coding algorithms. *Brain and Cognition*, *112*, 92–97.

Steriade, D. (1988). Review article: Clements and Keyser, CV phonology. *Language*, *64*, 118–129.

Stetson, R. H. (1951). *Motor phonetics* (2nd ed.). Amsterdam, The Netherlands: North-Holland.

Stilp, C. E., & Kluender, K. R. (2010). Cochlea-scaled entropy, not consonants, vowels, or time, best predicts speech intelligibility. *Proceedings of the National Academy of Sciences*, *107*(27), 12387–12392.

Stilp, C. E., & Kluender, K. R. (2016). Stimulus statistics change sounds from near-indiscriminable to hyperdiscriminable. *PLoS One*, *11*(8), e0161001. https://doi.org/10.1371/journal.pone.0161001

Stokoe, W. C. (1960). Sign language structure: An outline of the visual communication system of the American deaf. Studies in Linguistics: Occasional Papers 8.

Strickland, B., Geraci, C., Chemla, E., Schlenker, P., Kelepir, M., & Pfau, R. (2015). Event representations constrain the structure of language: Sign language as a window into universally accessible linguistic biases. *Proceedings of the National Academy of Sciences*, *112*(19), 5968–5973.

Swallow, K. M., Zacks, J. M., & Abrams, R. A. (2009). Event boundaries in perception affect memory encoding and updating. *Journal of Experimental Psychology: General*, *138*(2), 236–257.

Ten Oever, S., Hausfeld, L., Correia, J. M., Van Atteveldt, N., Formisano, E., & Sack, A. T. (2016). A 7T fMRI study investigating the influence of oscillatory phase on syllable representations. *NeuroImage*, *141*, 1–9. https://doi.org/10.1016/j.neuroimage.2016.07.011

ten Oever, S., & Sack, A. T. (2015). Oscillatory phase shapes syllable perception. *Proceedings of the National Academy of Sciences*, *112*(52), 15833–15837. https://doi.org/10.1073/pnas.1517519112

Theodore, R. M., Miller, J. L., & DeSteno, D. (2009). Individual talker differences in voice-onset-time: Contextual influences. *The Journal of the Acoustical Society of America*, *125*(6), 3974–3982. https://doi.org/10.1121/1.3106131

Vennemann, T. (1972). On the theory of syllabic phonology. *Linguistische Berichte*, *18*(1), 1–19.

Wilbur, R. B. (1979). *American sign language and sign systems*. New York, NY: University Park Press.

Wilbur, R. B. (2008). Complex predicates involving events, time and aspect: Is this why sign languages look so similar. In J. Quer (Ed.), *Signs of the time: Selected papers from Theoretical Issues in Sign Language Research (TISLR) 2004*, 217–250. Hamburg: Signum.

Wilbur, R. B. (2010). The semantics-phonology interface. In *Cambridge language surveys: Sign languages*. Cambridge: Cambridge University Press.

Wilbur, R. B. (2015). Word formationin sign languages. In P. O. Müller, I. Ohnheiser, S. Olsen, F. Rainer (Eds.), Word Formation. *An international handbook (HSK -Handbooks of linguistics and communication science)*, (pp. 2225–2251; chapter 126). Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110375732-037.

Wilbur, R. B., & Allen, G. D. (1991). Perceptual evidence against internal structure in American sign language syllables. *Language and Speech*, *34*(1), 27–46.

Wilbur, R. B., & Malaia, E. (2008). Event visibility hypothesis: Motion capture evidence for overt marking of telicity in ASL. Paper presented at Linguistic Society of America meeting, Chicago, IL. https://doi.org/10.3115/1626481.1626494

Wilbur, R. B., Malaia, E., & Shay, R. A. (2012). Degree modification and intensification in American sign language adjectives. In M. Aloni, V. Kimmelman, F. Roelofsen, G. Sassoon, K. Schulz & M. Westera (Eds.), *Logic, language and meaning*, 92–101. Berlin: Springer.

Wilbur, R. B., & Nolen, S. B. (1986). The duration of syllables in American sign language. *Language and Speech*, *29*(3), 263–280.

Wilbur, R. B., & Petersen, L. (1997). Backwards signing and ASL syllable structure. *Language and Speech*, *40*(1), 63–90.

Wilbur, R. B., & Schick, B. S. (1987). The effects of linguistic stress on ASL signs. *Language and Speech*, *30*(4), 301–323.

Woodruff Carr, K., White-Schwoch, T., Tierney, A. T., Strait, D. L., & Kraus, N. (2014). Beat synchronization predicts neural speech encoding and reading readiness in preschoolers. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(40), 14559–14564. https://doi.org/10.1073/pnas.1406219111

Zacks, J. M., Braver, T. S., Sheridan, M. A., Donaldson, D. I., Snyder, A. Z., Ollinger, J. M., … Raichle, M. E. (2001). Human brain activity time-locked to perceptual event boundaries. *Nature Neuroscience*, *4*(6), 651–655.

Zacks, J. M., Kumar, S., Abrams, R. A., & Mehta, R. (2009). Using movement and intentions to understand human activity. *Cognition*, *112*(2), 201–216.

Zacks, J. M., Swallow, K. M., Vettel, J. M., & McAvoy, M. P. (2006). Visual motion and the neural correlates of event perception. *Brain Research*, *1076*(1), 150–162. https://doi.org/10.1016/j.brainres.2005.12.122

Zacks, J. M., & Tversky, B. (2001). Event structure in perception and conception. *Psychological Bulletin*, *127*(1), 3–21.

Zoefel, B., Archer-Boyd, A., & Davis, M. H. (2018). Phase entrainment of brain oscillations causally modulates neural responses to intelligible speech. *Current Biology*, *28*(3), 401–408.e5. https://doi.org/10.1016/j.cub.2017.11.071