The syllable in the light of motor skills and neural oscillations

Antje Strauss, Jean-Luc Schwartz

HAL Id: hal-01424458
https://hal.science/hal-01424458

Submitted on 2 Jan 2017

Title: The syllable in the light of motor skills and neural oscillations

Authors: Antje Strauß[1,2, *], & Jean-Luc Schwartz[1]

Affiliations:

[1]Univ. Grenoble Alpes, CNRS, GIPSA-lab, F-38000 Grenoble, France

[2]Zukunftskolleg & Department of Linguistics, University of Konstanz, Germany

Running Title: Syllables and theta oscillations

*corresponding author:

Antje Strauß

Universitätsstr. 10

78457 Konstanz

antje.strauss@uni-konstanz.de

Abstract

Recent advances in neuroscience have brought a great focus on how the auditory cortex tracks speech at certain time scales corresponding to pre-lexical speech units in order to achieve comprehension. In particular, it has been claimed that it is the syllabic rhythm to which slow neural oscillations in the auditory cortex entrain in order to chunk the speech stream into smaller informational units. However, the terms "syllable" and "rhythm" have been treated quite loosely in the current literature. We revisit classic approaches to show that both concepts do not necessarily have an acoustic or phonetic counterpart, which could be directly extracted by neural processes. We would like to suggest that the syllabic rhythm could emerge at the intersection of acoustic–phonetic and motor knowledge of speech. We furthermore propose that nesting of cortical oscillations might be the key mechanism to understand the timing constraints that lead to the emergence of the syllable.

Keywords: Neural oscillations; syllabic rhythm; perceptuo-motor integration; envelope entrainment; temporal prediction

Introduction

Recent advances in neuroscience have brought a great focus on how the auditory cortex tracks speech at certain time scales corresponding to pre-lexical speech units (phonemes and syllables mainly) in order to achieve comprehension. In particular, it has been claimed that it is the syllabic rhythm to which slow neural oscillations (3–7 Hz) in the auditory cortex entrain in order to chunk the speech stream into smaller informational units (Giraud & Poeppel, 2012). However, as we argue in this paper, the terms "syllable" and "rhythm" have been treated quite loosely in the current literature. Both concepts do not necessarily have an acoustic or phonetic counterpart, which could be directly extracted by neural processes. But we would like to suggest here that a syllabic rhythm could emerge by considering also the motor knowledge of speech production that gives important clues on the temporal predictability in connected speech. Therefore, we first elaborate on the complex relationships between linguistic syllables and acoustic/ phonetic information on the one hand and articulatory motor rhythms on the other hand. Then, we outline a framework for how to use these relationships in order to determine the functional role of slow neural oscillations in auditory cortex. Finally, we propose that nesting of cortical oscillations might be the key mechanism to understand the link between acoustic, perceptual, linguistic and motor constraints that lead to the emergence of the syllable.

Defining the syllable and the syllabic rhythm

Linguistic, acoustic-phonetic and psycholinguistic approaches to the syllable. The syllable plays a central role in most phonological and psycholinguistic theories of speech recognition. The classical linguistic approach to describe how a syllable is organized is based on the syntagmatic relationships between phonemes (see Fig. 1A). It disassembles the syllable usually into three slots namely onset, nucleus and coda where onset and coda consist of zero, one or several consonants and where the nucleus consists obligatorily of at least one vowel (though some languages like Slovak accept syllabic consonants as the nucleus) (Hockett, 1955).

The classical phonetic description of the syllable is based on the waxing and waning of sonority (Bloch & Trager, 1942) that is the relative amount of energy used for producing speech. Thus, it has been hypothesised that within a syllable phonemes are sorted from the least sonorous like stop consonants, followed by fricatives, nasals, liquids, glides to vowels as the most sonorous segments at the nuclear syllable position, and are conversely ordered in the coda if there is one. Whereas some authors confirm the correspondence between sonority fluctuations and syntagmatically defined syllables (as it is depicted in Fig. 1B; Bloomfield, 1933), others doubt that there is a phonetic reality of the syllable (Haugen, 1956) and might concede sonority fluctuations to correspond to simple CV-sequences but not to more complex syllables (Bloomfield, 1933; Ohala, 2008; Cummins, 2012). In fact, there are many exceptions to the sonority rule in all languages. Very common examples are cases of fricative–plosive-onsets in syllables like /spa/ where /s/ precedes /p/ although /p/ is categorized as less sonorous than /s/. Others (including the recent neuroscientific literature) have tried to link local maxima in the amplitude modulation of the speech envelope to the syllable. Here, monosyllables like /spa/ would again constitute an exception because of their two amplitude peaks separated by the silent period associated with the closure of the /p/.
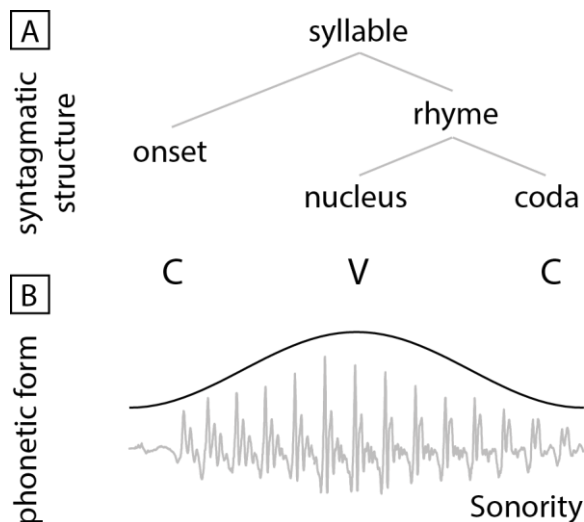

Figure 1: A. Syllable structure in the classical linguistic literature. **B.** Syllable structure as an acoustic-phonetic sequence of increasing-decreasing sonority. C = consonant, V = vowel.

The legitimate existence of the psycholinguistic syllable was originally proven in a pivotal study by Mehler, Dommergues, Frauenfelder, and Segui (1981). They had shown that phoneme sequences were easier to detect if they correspond to a word's syllabification. In their case, the CV-sequence "ba" was detected faster in "ba-lance" than in "bal-con", whereas the CVC-sequence "bal" was detected faster in "bal-con" than in "ba-lance". They concluded that the syllable holds a central role in the perceptual organization of the speech stream, and declared the syllable as a pivotal unit of speech perception.

However, replications of this effect turned out to be difficult (Cutler, 1997; Cutler, Mehler, Norris, & Segui, 1983, 1986) and might depend on the rhythmic class of a language (e.g., Cutler, McQueen, Norris, & Somejuan, 2001). In the same vein, syllabification is often ambiguous (Content, Kearns, & Frauenfelder, 2001; Treimann & Danis, 1988) and even the existence of ambisyllabic consonants (consonants belonging to two syllables at the same time) has been claimed (Kahn, 1980). Thus finally, serial models where information is sent in discrete chunks like syllables have been discarded in the psycholinguistic community (Cutler et al., 2001). It has rather been argued that the process of speech segmentation and lexical access is based on statistical processes including syllabic structure as one of several cues operating jointly to enable the listener to achieve adequate word segmentation (Norris, McQueen, Cutler, & Butterfield, 1997; Norris & McQueen, 2008).

Syllabic rhythm and its perception. Rhythmic analyses of speech have determined an average syllable length of about 200 to 250 ms resulting in a 4 to 5 Hz rhythm (Ohala, 1975; Chandrasekaran, Trubanova, Stillittano, Caplier, & Ghazanfar, 2009). More detailed efforts to define the syllable rhythm acoustically show quite some variation to this average. For example, in some languages longer (stressed) syllables are frequently surrounded by much shorter (unstressed) syllables. If, however, syllables approach the average duration of 250 ms, the probability increases that neighbouring syllables will be also around the same length, hence creating some sort of isochronicity (Greenberg, 1999; Greenberg, Carvey, Hitchcock, & Chang, 2003).

A sensitive measure to characterize syllabic rhythm changes over time in a single utterance have turned out to be difficult. If successful it might finally lead to an objective classification of the world's languages into stress-timed (like English or German), syllable-timed (like

French or Italian) or mora-timed (like Japanese) languages (Ramus, Nespor, & Mehler, 1999). Typically, a ratio is calculated in a given utterance between the percentage of vowel durations and the standard deviation of consonant durations. But although reproducing the traditional classification, it is in fact influenced by the tendency that "stress-timed" languages use more complex consonant clusters compared to "syllable-timed" languages. Thus, it appears to be less meaningful concerning the rhythmic changes across time and not to be robust against changes in articulation rate. The normalized syllable duration "Yet Another Rhythm Determination" proposed by Wagner and Dellwo (2004) might provide a possibility to capture the rhythmic irregularities within an utterance that accounts for the greater variability across adjacent syllables in stress-timed languages.

Despite the problem how to quantify the syllabic rhythm, another branch of research asks what the basis of speech rhythm perception is. Some authors showed that amplitude modulations at ~4 Hz but not at faster rates contribute to speech rhythm perception (Leong, Stone, Turner, & Goswami, 2014). But it might be that listeners do not consider syllable onsets for their rhythm judgments but rather the distances between p-centres of syllables (Morton, Marcus, & Frankish, 1976; Marcus, 1981; Scott, 1998). The p-centre (i.e., the perceptual centre) is defined as the "psychological moment of occurrence". The precise physical correlate of p-centres, though, is still under debate and most likely describes the conjunction between multiple (e.g., acoustic and articulatory) cues (DeJong, 1994; Patel, Löfqvist, & Naito, 1999).

The speech motor syllable and motor knowledge of syllable timing. The syllable is considered to play a central role in speech production as well. Speech errors like slips-of-the-tongue led Crompton (1981) to postulate the existence of a mental syllabary, which was later seized by Levelt (1999) as how phonological commands are translated into motor gestures.

In terms of motor activity, the syllable has been described as the relative opening and closing of the vocal tract (corresponding to sonority fluctuations; Goldsmith, 1990). The motor syllable holds a central role in the Frame-Content Theory (MacNeilage, 1998; MacNeilage & Davis, 2000) where it is seen as the "frame" allowing further phonological "content". It is at around 7 months of age that infants start to babble by alternating between close and open mouth configurations (Oller, 1980), the so-called "closants" and "vocants"

(Martin, 1981). While indeed vocalizations precede articulation in Human development as well as in Humanoid evolution (Oller & Griebel, 2008), speech similar utterances begin as soon as infants are able to couple vocalizations with orofacial gestures that is when they start to babble. At this first stage of development, utterances are thus dominated by the cyclic movements of the mandible (Green, Moore, Higashikawa, & Steeve, 2000). The development of independent phonemic content inside the frame would only occur at a later stage.

The phylogenetical component of the Frame-Content Theory further hypothesizes that syllables evolved from mastication and other ingestive gestures. Indeed, jaw cycles in speech, mouth openings and syllable rate are highly correlated (Chandrasekaran et al., 2009). However, jaw cycles in chewing are much slower in frequency (Hiiemae et al., 2002). This is one argument to favour the faster, communicative lip smacking as observed in monkeys as the precursor of the syllable (Ghazanfar, Takahashi, Mathur, & Fitch, 2012).

Whatever the view on development and phylogeny, it is true that in adult speech, the syllable involves complex sequences of articulatory gestures in which the jaw is only, at best, a carrier articulator. Among the concurrent perspectives on the motor structure of syllables, one of the most sophisticated models is provided by articulatory phonology (Browman & Goldstein, 1986), which tries to explain phonological structures by combinatorial constraints of articulatory gestures. The articulatory coordination between consonants and vowels within a syllable is framed as coupled and decoupled oscillators. That means, consonants at the onset are produced "in phase" with vowels at the nucleus because their articulation is prepared in parallel whereas the production of the nuclear vowels and consonants at the coda are somewhat decoupled (Browman & Goldstein, 1988; Goldstein, Nam, Saltzman, & Chitoran, 2009). Interestingly, it is furthermore suggested that the temporal alignment of onset consonants and nuclear vowels is constant. Hence especially when the syllable onset consists of a complex consonant cluster, articulation is supposed to be programmed around a "c-centre" (i.e., the consonantal centre), a sort of mean of consonantal targets in the onset, in order to keep a constant temporal distance to the vowel centre.

In summary, the coupled articulatory coordination at syllable onsets might have neural correlates of coupled oscillations in motor cortex. Together with the quasi-regular

oscillation of the jaw, they might form a good basis for motor knowledge of syllable timing, which might be also of use during speech perception (Wilson & Wilson, 2005; Wilson, Saygin, Sereno, & Iacoboni, 2004) especially in adverse listening conditions, as we will outline in the next section.

The emergence of the syllable at the linguistic, acoustic and motor intersection

As we have seen, the definition of the syllable and its role in speech comprehension remains challenging for linguists, psycholinguists and phoneticians. In neurolinguistics, the syllable has not played an important role up to very recently. Some researchers have proposed modular processes (Friederici, 2002) and others cascaded processes (Hagoort, 2008) to achieve lexical access without being too detailed about pre-lexical units like syllables. An interesting perspective has arrived based on neural oscillations, which sheds a new (syllable-based) light onto this debate. Several studies have by now shown that slow neural oscillations in the auditory cortex track the envelope of the speech signal (Ahissar et al., 2001; Luo & Poeppel, 2007; for a review see Ding & Simon, 2014). The suggested mechanism behind is that intrinsic neural oscillations in the theta frequency band (3–7 Hz) would adapt in frequency to extrinsic acoustic rhythms at a syllabic rate (~4 Hz; Ghitza, 2011) so that microstates of high cortical excitability are aligned with the most informative parts of the speech signal (Giraud & Poeppel, 2012). As we have outlined above, the relationship of energy fluctuations, which constitute the speech envelope to a great extent, and syllable structure is not as straight forward. The same holds true for the relationship between envelope and rhythmic perception. Nevertheless, envelope entrainment has often been referred to as tracking the syllable rhythm in the neuroscientific literature, but we are far from having a complete picture of the theta-mediated cognitive mechanisms and in particular of its relation to the syllable. In the following, we will first elaborate on the role of theta oscillations when decoding the speech signal acoustically, and second, we provide new ideas about how theta oscillations may be the common ground of how motor rhythms interact with acoustic decoding.

Theta oscillations in primary auditory cortex and pre-lexical abstraction. As outlined above, current psycholinguistic models on word recognition postulate the existence of pre-lexical units for the efficient mapping of sound onto meaning. In neuroscience, efforts to localize the representation of pre-lexical units have been made and suggest superior temporal sulcus and gyrus as the main loci (for review see Obleser & Eisner, 2009). However, the nature of these pre-lexical units remains unclear. In all psycholinguistic models, the pre-lexical abstraction process has been simplified for the sake of modelling higher-level aspects of the word recognition process (for review see Scharenborg, Norris, ten Bosch, & McQueen, 2005). In Shortlist B, for example, ready-made phonemes are given as an input to the model, and constitute the basis for calculating higher-level probabilities of upcoming phonemes and of the final word segmentation using Bayesian principles.

Interestingly, the recent neurophysiological model by Giraud & Poeppel (2012) might provide a way, how abstract pre-lexical, that is phonemic and syllabic, chunks are formed in the first place. They propose that gamma oscillations in primary auditory cortex decode acoustic-phonetic information in a spatio-temporal manner. This gamma code, however, needs to be integrated similar to second level statistics, which is according to them provided by oscillations in the theta frequency range. Furthermore, gamma-theta coupling might be implemented across different cortical layers of primary auditory cortex. Most importantly, theta oscillations hence mark a window of temporal integration that resembles in its duration linguistically defined syllables.

On the basis of this temporal resemblance, some authors have made strong deterministic claims about how theta oscillations chunk speech streams into syllable-sized units (Ghitza, 2011). The functional role of theta oscillations, however, might neither be syllabification nor speech segmentation. In contrast to the claims raised by linguistic approaches to find a final definition of the syllable (by its syntagmatic structure and its sonority fluctuations), the neurophysiological model of gamma-theta coupling instead might provide a less deterministic and more probabilistic approach to the syllable. Here, theta oscillations might only modulate the pre-lexical abstraction process but not chunk the speech stream across time. This notion is supported by recent results showing that theta phase influences phonemic categorization (Ten Oevers & Sack, 2015; Ten Oevers et al., 2016), whereas its role in speech segmentation could not yet be established (Kösem, Basirat, Azizi, & Van

Wassenhove, 2016; Strauß et al., in prep.). This would mean that the acoustic information is decoded in a phase-dependent manner in order to form abstract phonemic representations, but this might be important to ensure a robust representation against noise and hence is unrelated to speech segmentation (Kayser, Montemurro, Logothetis, & Panzeri, 2009). Such a probabilistic modulation would furthermore be in line with the current opinion in psycholinguistics, where discrete syllabic chunks have been discarded in favour of a cascaded accumulation of information (Cutler et al., 2001).

Theta oscillations and the missing link to motor cortex. Current models on the functional role of theta oscillations focus primarily on the parsing of the speech signal (Giraud & Poeppel, 2012; Ghitza, 2013). However, multisensory influences like visual input have been shown recently to modify speech envelope tracking and in particular theta oscillations in primary auditory cortex (Crosse, Butler, & Lalor, 2015; Zion Golumbic, Cogan, Schroeder, & Poeppel, 2013; for review see Peelle & Sommers, 2015). In the light of the current approach, it is noteworthy that fluctuations in the speech envelope or in energy are also associated with the opening of articulators, in particular with the jaw, to produce vowels and the closing of articulators to produce consonants. Thus, if we pose the evolutionary question whether spoken language tries to match intrinsic cortical oscillations (or the other way round), we inevitably need to answer the question whether there is a specific relationship of the auditory theta rhythm to motor cortex and to articulatory skills.

In order to study this specific relationship, we suggest a tentative schema about potential interactions following from the previous elaborations (see Figure 2). In short, we propose that syllables emerge from a complex set of interactions that comprise not only speech acoustics and slow neural oscillations but also motor rhythms.
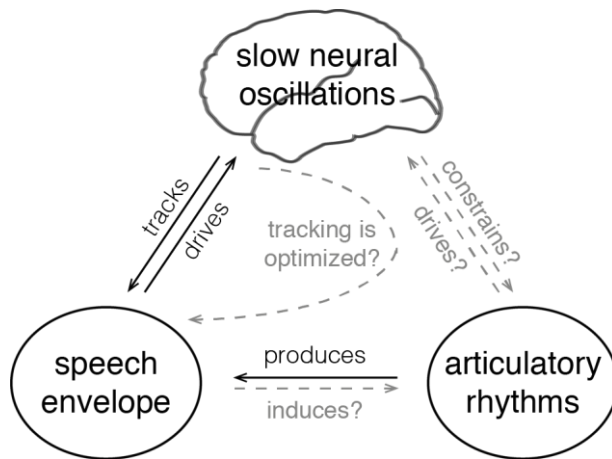
Figure 2: Schema of putative relationships between acoustic rhythms, neural oscillations and motor knowledge letting syllables emerge from this set of interactions. Solid arrows mark relationships that are reasonably well accepted; dotted arrows mark relationships that could be object of future studies and developments in the field.

Concerning the first link, namely the relationship between slow neural oscillations in auditory cortex and the speech envelope, we have already reviewed several studies above that have targeted this question. Some might say that the speech envelope drives theta oscillations in primary auditory cortex whereas others would say that theta oscillations track speech envelope fluctuations (hence rather solid lines for both processes in Fig. 2). Secondly, we depict (also with a solid line in Fig. 2) the link between articulatory movements that produce the speech envelope as discussed in previous sections. Motor rhythms might encompass simple jaw oscillations up to sophisticated coupling between jaw movements and all the other orofacial articulators in order to produce "the smallest unit of speech utterances", that is the syllable (Ladefoged & Johnson, 2011).

Links that are more questionable and programmatic are drawn in dotted lines in Figure 2. On the one hand, there is the most important question whether slow neural oscillations constrain articulatory rhythms. In their Frame-Content theory, MacNeilage & Davis (2000) suggested that jaw rhythms provided the basis for speech rhythms. Hence, natural rhythms of the speech envelope would be constrained by natural motor rhythms. However, the current neuroscientific framework suggests that the speech envelope accommodates theta oscillations pre-existing in auditory cortex. Theta oscillations might therefore also provide a

"neural temporal frame" for motor oscillations (e.g. Giraud, Kleinschmidt, Poeppel, Lund, & Laufs, 2007; Morillon et al., 2010). Considering that jaw oscillations in e.g. chewing seem to be intrinsically slower than speech rhythms (Ghazanfar et al., 2012), it could be argued that the theta frame indeed induced a speeding of jaw rhythms aiming at better fitting with the auditory decoding capacities of the human brain (Ghitza, 2011).

On the other hand, a considerable number of studies have shown that motor cortices probably contribute to auditory perception particularly in adverse conditions (e.g. Meister et al., 2007; D'Ausilio et al., 2009; Sato, Tremblay, & Gracco, 2009; for meta-analyses and reviews see Scott, McGettigan, & Eisner, 2009; Adank, 2012). Its role for speech processing is still open, but some evidence suggests that the motor cortex might monitor rhythm and rate in auditory signals (probably associated with articulatory rhythms) in order to improve temporal predictions of auditory targets (Arnal, Doelling, & Poeppel, 2015; Morillon, Schroeder, & Wyart, 2014). Interestingly, first evidence shows that the motor-related beta rhythm is used for predictive timing of auditory events (Fujioka, Trainor, Large, & Ross, 2012; for review see Arnal, 2012). Hence, slow oscillations in auditory cortex might couple with beta oscillations in order to increase the accuracy of temporal predictions (Arnal et al., 2015). This "when"-path has been suggested to be implemented in the dorsal stream (Rauschecker & Scott, 2009; Rauschecker, 2011). If such auditory-motor coupling could be established, one could further ask whether reported visual enhancement of speech envelope tracking is actually mediated by motor cortices (as there is first evidence pointing into this direction by Park, Kayser, Thut, & Gross, 2016). This would be in line with some aspects of the classic motor theory of speech perception (Liberman & Mattingly, 1985). In any case in speech, such temporal predictions would be indeed helpful in order to anticipate suprasegmental time windows of information integration like phrase boundaries (Ding, Melloni, Tian, Zhang, & Poeppel, 2015) or to arrive at an efficient turn taking in a conversation (Wilson & Wilson, 2005; Scott et al., 2009).

In sum with our model, we postulate intimate connections between auditory (and possible visual), motor and neural processes inside the "syllabic" temporal frame. The tracking of acoustic fluctuations over time could be optimized thanks to the recruitment of articulatory knowledge, e.g. via the dorsal stream, especially in adverse listening conditions. This would

constitute an efficient system for integrating information in time even under degraded conditions of communication.

Conclusion

It is now becoming increasingly clear that speech communication involves a set of tight relationships between the perceptual and the motor system, in both speech production (Guenther, 2006; Perrier, 2005) and speech perception (Schwartz, Basirat, Ménard, & Sato, 2012; Skipper, van Wassenhove, Nusbaum, & Small, 2007). However, computational models for integrating motor and perceptual processes into one cognitive speech communication systems remain rather formal and far from detailed neural implementation (Guenther, Hampson, & Johnson, 1998; Moulin-Frier, Laurent, Bessière, Schwartz, & Diard, 2012). In this paper, we suggest that neural oscillations could provide a natural framework for perceptuo-motor integration in the emergence of syllabic units in the listener's brain. This opens a large number of experimental questions and theoretical perspectives for both speech perception, speech production and the description of the pre-lexical units of spoken languages.

Acknowledgements

Funding

References

Adank, P. (2012). The neural bases of speech comprehension and speech production: Two activation Likelihood Estimation (ALE) meta-analyses. *Brain and Language,* 22(1), 42–54. DOI: 10.1016/j.bandl.2012.04.014

Ahissar, E., Nagarajan, S., Ahissar, M., Protopapas, A., Mahncke, H., & Merzenich, M.M. (2001). Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. *Proceedings of the National Academy of Sciences of the United States of America,* 98(23), 13367–13372. DOI: 10.1073/pnas.201400998

Arnal, L. (2012). Predicting "when" using the motor system's beta-band oscillations. *Frontiers in Human Neuroscience,* 6, 225. DOI: 10.3389/fnhum.2012.00225

Arnal, L., Doelling, K., & Poeppel, D. (2015). Delta–Beta Coupled Oscillations Underlie Temporal Prediction Accuracy. *Cerebral Cortex,* 25(9), 3077–3085. DOI: 10.1093/cercor/bhu103

Bloch, B., & Trager, G. (1942). Outlines of Linguistic Analysis. Baltimore, MD: Linguistic Society of America.

Bloomfield, L. (1933). Language. Chicago: University of Chicago Press (reprint of 1984).

Browman, C.P., & Goldstein, L. (1986). Towards an articulatory phonology. *Phonology Yearbook,* 3, 219–252.

Browman, C.P., & Goldstein, L. (1988). Some notes on syllable structure in Articulatory Phonology. *Phonetica,* 45(2–4), 140–155.

Chandrasekaran, C., Trubanova, A., Stillittano, S., Caplier, A., & Ghazanfar, A.A. (2009). The natural statistics of audiovisual speech. *PLoS Computational Biology,* 5(7), e1000436. DOI: 10.1371/journal.pcbi.1000436

Content, A., Kearns, R.K., & Frauenfelder, U.H. (2001). Boundaries versus onsets in syllabic segmentation. *Journal of Memory and Language,* 45(2), 177–199. DOI: 10.1006/jmla.2000.2775

Crompton, A. (1981). Syllables and segments in speech production. *Linguistics,* 19(7– 8), 663–716.

Crosse, M.J., Butler, J.S., & Lalor, E.C. (2015). Congruent visual speech enhances cortical entrainment to continuous auditory speech in noise-free conditions. *The Journal of Neuroscience,* 35(42), 14195–14204. DOI: 10.1523/JNEUROSCI.1829-15.2015

Cummins, F. (2012). Oscillators and syllables: a cautionary note. *Frontiers in Psychology,* 3, 364. DOI: 10.3389/fpsyg.2012.00364a

Cutler, A. (1997). The syllable's role in the segmentation of stress languages. *Language and Cognitive Processes,* 12, 839–845. DOI: 10.1080/016909697386718

Cutler, A., McQueen, J.M., Norris, D., & Somejuan, A. (2001). The Roll of the Silly Ball. In E. Dupoux (Ed.), *Language, Brain, and Cognitive Development* (pp. 181–194). Cambridge: MIT Press.

Cutler, A., Mehler, J., Norris, D., & Segui, J. (1983). A language specific comprehension strategy. *Nature,* 304, 159–160.

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

D'Ausilio, A., Pulvermüller, F., Salmas, P., Bufalari, I., Begliomini, C., & Fadiga, L. (2009). The motor somatotopy of speech perception. *Current Biology*, 19, 381–385.

DeJong, K. (1994). The correlation of P-center adjustments with articulatory and acoustic events. *Perception & Psychophysics,* 56(4), 447–460. DOI: 10.3758/BF03206736

Ding, N., Melloni, L., Tian, X., Zhang, H., & Poeppel, D. (2015). Cortical tracking of hierarchical linguistic structures in connected speech. *Nature Neuroscience,* 19, 158–164. DOI: 10.1038/nn.4186

Ding, N., & Simon, J.Z. (2014). Cortical Entrainment to Continuous Speech: Functional Roles and Interpretations. *Frontiers Human Neuroscience,* 8, 311. DOI: 10.3389/fnhum.2014. 00311

Friederici, A.D. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences,* 6(2), 78–84. DOI:10.1016/s1364-6613(00)01839-8

Fujioka, T., Trainor, L.J., Large, E.W., & Ross, B. (2012). Internalized timing of isochronous sounds is represented in neuromagnetic β oscillations. *The Journal of Neuroscience,* 32(5), 1791–1802. DOI: 10.1523/JNEUROSCI.4107-11.2012

Ghazanfar, A.A., Takahashi, D.Y., Mathur, N., & Fitch, W.T. (2012). Cineradiography of monkey lipsmacking reveals the putative origins of speech dynamics. *Current Biology,* 22(13), 1176–1182. DOI: 10.1016/j.cub.2012.04.055

Ghitza, O. (2011). Linking speech perception and neurophysiology: speech decoding guided by cascaded oscillators locked to the input rhythm. *Frontiers in Psychology,* 2, 130. DOI: 10.3389/fpsyg.2011.00130

Ghitza, O. (2013). The theta-syllable: a unit of speech information defined by cortical function. *Frontiers in Psychology,* 4, 138. DOI: 10.3389/fpsyg.2013.00138

Giraud, A.L., Kleinschmidt, A., Poeppel, D., Lund, T.E., & Laufs, H. (2007). Endogenous cortical rhythms determine cerebral specialization for speech perception and production. *Neuron,* 56(6), 1127–1134. DOI: 10.1016/j.neuron.2007.09.038

Giraud, A.L., & Poeppel, D. (2012). Cortical oscillations and speech processing: emerging computational principles and operations. *Nature Neuroscience,* 15(4), 511–517. DOI: 10.1038/nn.3063

Goldsmith, J.A. (1990). Autosegmental and metrical phonology. Oxford: Blackwell.

Goldstein, L., Nam, H., Saltzman, E., & Chitoran, I. (2009). Coupled oscillator planning model of speech timing and syllable structure. In G. Fant, H. Fujisaki, & J. Shen (Eds.), *Frontiers in phonetics and speech science* (pp. 239-250). Beijing: The Commercial Press.

Green, J.R., Moore, C.A., Higashikawa, M., & Steeve, R.W. (2000). The Physiologic Development of Speech Motor Control: Lip and Jaw Coordination. *Journal of Speech, Language, and Hearing Research,* 43(1), 239–255. DOI: 10.1044/jslhr.4301.239

Greenberg, S. (1999). Speaking in shorthand – A syllable-centric perspective for understanding pronunciation variation. *Speech Communication,* 29(2–4), 159–176. DOI: 10.1016/S0167-6393(99)00050-3

Greenberg, S., Carvey, H., Hitchcock, L., & Chang, S. (2003). Temporal properties of spontaneous speech—a syllable-centric perspective. *Journal of Phonetics,* 31(3–4), 465–485. DOI: 10.1016/j.wocn.2003.09.005

Guenther, F.H. (2006). Cortical interactions underlying the production of speech sounds. *Journal of Communication Disorders,* 39(5), 350–365. DOI: 10.1016/j.jcomdis.2006.06.013

Guenther, F.H., Hampson, M., & Johnson, D. (1998). A theoretical investigation of reference frames for the planning of speech movements. *Psychological Review,* 105(4), 611–633. DOI: 10.1037//0033-295X.105.4.611-633

Hagoort, P. (2008). The fractionation of spoken language understanding by measuring electrical and magnetic brain signals. *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences,* 363(1493), 1055–1069. DOI: 10.1098/rstb.2007.2159

Haugen, E. (1956). The syllable in linguistic description. In M. Halle, H.G. Lunt, H. McClean (Eds.), *For Roman Jakobson* (pp. 213–221). The Hague: Mouton.

Hiiemae, K.M., Palmer, J.B., Medicis, S.W., Hegener, J., Jackson, B.S., & Lieberman, D.E. (2002). Hyoid and tongue surface movements in speaking and eating. *Archives of Oral Biology,* 47(1), 11–27. DOI: 10.1016/S0003-9969(01)00092-9

Hockett, C.F. (1955). *A manual of phonology*. Baltimore: Waverly Press.

Kahn, D. (1980). *Syllable-based Generalizations in English Phonology*. New York: Garland Publishing.

Kayser, C., Montemurro, M.A., Logothetis, N.K., & Panzeri, S. (2009). Spike-Phase Coding Boosts and Stabilizes Information Carried by Spatial and Temporal Spike Patterns. *Neuron,* 61, 597–608. DOI: 10.1016/j.neuron.2009.01.008

Kösem, A., Basirat, A., Azizi, L., & Van Wassenhove, V. (2016). High frequency neural activity predicts word parsing in ambiguous speech streams. *Journal of Neurophysiology.* DOI: 10.1152/jn.00074.2016

Ladefoged, P., & Johnson, K. (2011). *A course in phonetics* (6th ed., pp. 243–266) Boston, MA: Wadsworth.

Leong, V., Stone, M.A., Turner, R.E., & Goswami, U. (2014). A role for amplitude modulation phase relationships in speech rhythm perception. *The Journal of the Acoustical Society of America,* 136(1), 366–381. DOI: 10.1121/1.4883366

Levelt, W.J.M. (1999). Models of word production. *Trends in Cognitive Sciences,* 3(6), 223–232. DOI: 10.1016/S1364-6613(99)01319-4

Liberman, A.M., & Mattingly, I.G. (1985). The motor theory of speech perception revised. *Cognition,* 21(1), 1–36. DOI: 10.1016/0010-0277(85)90021-6

Luo, H., & Poeppel, D. (2007). Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron,* 54(6), 1001–1010. DOI: 10.1016/j.neuron.2007.06.004

MacNeilage, P., & Davis, B.L. (2000). On the origin of internal structure of word forms. *Science,* 288(5465), 527–531. DOI: 10.1126/science.288.5465.527

MacNeilage, P.F. (1998). The frame/content theory of evolution of speech production. *Behavioral and Brain Sciences,* 21(04), 499–546.

Marcus, S.M. (1981). Acoustic determinants of perceptualcenter (P-center) location. *Perception & Psychophysics,* 30(3), 247–256. DOI: 10.3758/BF03214280

Martin, J.A.M. (1981). *Voice, speech and language in the child: Development and disorder.* New York: Springer.

Mehler, J., Dommergues, J.Y., Frauenfelder, U., & Segui, J. (1981). The syllable's role in speech segmentation. *Journal of Verbal Learning and Verbal Behavior,* 20(3), 298–305. DOI: 10.1016/S0022-5371(81)90450-3

Meister, I. G., Wilson, S. M., Deblieck, C., Wu, A. D., & Iacoboni, M. (2007). The essential role of premotor cortex in speech perception. *Current Biology*, 17, 1692–1696.

Morillon, B., Lehongre, K., Frackowiak, R. S. J., Ducorps, A., Kleinschmidt, A., Poeppel, D., & Giraud, A. L. (2010). Neurophysiological origin of human brain asymmetry for speech and language. *Proceedings of the National Academy of Sciences of the United States of America*, 107(43), 18688– 18693.

Morillon, B., Schroeder, C.E., & Wyart, V. (2014). Motor contributions to the temporal precision of auditory attention. *Nature Communications,* 5, 5255. DOI: 10.1038/ncomms6255

Morton, J., Marcus, S.M., & Frankish, C. (1976). Perceptual Centers (P-centers). *Psychological Review,* 83(5), 405–408.

Moulin-Frier, C., Laurent, R., Bessière, P., Schwartz, J.L., & Diard, J. (2012). Adverse conditions improve distinguishability of auditory, motor, and perceptuo-motor theories of speech perception: An exploratory Bayesian modelling study. *Language and Cognitive Processes,* 27(7–8), 1240–1263. DOI: 10.1080/01690965.2011.645313

Norris, D., & McQueen, J.M. (2008). Shortlist B: A Bayesian model of continuous speech recognition. *Psychological Review,* 115(2), 357–395. DOI: 10.1037/0033-295X.115.2.357

Norris, D., McQueen, J.M., Cutler, A., & Butterfield, S. (1997). The possible word constraint in the segmentation of continuous speech. *Cognitive Psychology,* 34(3), 191-243. DOI: 10.1006/cogp.1997.0671

Obleser, J., & Eisner, F. (2009). Pre-lexical abstraction of speech in the auditory cortex. *Trends in Cognitive Sciences,* 13(1), 14–19. DOI: 10.1016/j.tics.2008.09.005

Ohala, J.J. (1975). The Temporal Regulation of Speech. In G. Fant, & M.A.A. Tatham (Eds.), *Auditory Analysis and Perception of Speech* (pp. 431–453). London: Academic Press.

Ohala, J.J. (2008). The Emergent Syllable. In B.L. Davis & K. Zajdó (Eds.), *The Syllable in Speech Production* (pp. 179–186). New York: Lawrence Erlbaum Associates.

Oller, D.K. (1980). The emergence of speech sounds in infancy. In G. Yeni-Komshian, J.F. Kavanagh, & G.A. Ferguson (Eds.), *Child Phonology, vol 1. Production* (pp. 93–112). New York: Academic Press.

Oller, D.K., & Griebel, U. (2008). The Origins in Syllabification in Human Infancy and in Human Evolution. In B.L. Davis & K. Zajdó (Eds.), *The Syllable in Speech Production* (pp. 29–62). New York: Lawrence Erlbaum Associates.

Park, H., Kayser, C., Thut, G., & Gross, J. (2016). Lip movements entrain the observers' low-frequency brain oscillations to facilitate speech intelligibility. *eLife,* 5, e14521. DOI: 10.7554/eLife.14521

Patel, A.D., Löfqvist, A., & Naito, W. (1999). The acoustics and kinematics of regularly timed speech: A database and method for the study of the p-center problem. *Proceedings of the XIVth International Congress of Phonetic Sciences*, 1, 405–408.

Peelle, J. E., & Sommers, M. S. (2015). Prediction and constraint in audiovisual speech perception. *Cortex*, 68, 169–181.

Perrier, P. (2005). Control and representations in speech production. *ZAS Papers in Linguistics,* 40, 109–132.

Ramus, F., Nespor, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition,* 73(3), 265–292. DOI: 10.1016/S0010-0277(00)00101-3

Rauschecker, J.P. (2011). An expanded role for the dorsal auditory pathway in sensorimotor integration and control. *Hearing Research,* 271(1–2), 16–25. DOI: 10.1016/j.heares.2010.09.001

Rauschecker, J.P., & Scott, S.K. (2009). Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nature Neuroscience,* 12(6), 718–725. DOI: 10.1038/nn.2331

Sato, M., Tremblay, P., & Gracco, V. L. (2009). A mediating role of the premotor cortex in phoneme segmentation. *Brain and Language*, 111(1), 1–7. doi:10.1016/j.bandl.2009.03.002

Scharenborg, O., Norris, D., ten Bosch, L., & McQueen, J.M. (2005). How should a speech recognizer work? *Cognitive Science,* 29, 867–918. DOI: 10.1207/s15516709cog0000_37

Schwartz, J.L., Basirat, A., Ménard, L., & Sato, M. (2012). The Perception-for-Action-Control Theory (PACT): A perceptuo-motor theory of speech perception. *Journal of Neurolinguistics,* 25(5), 336–354. DOI: 10.1016/j.jneuroling.2009.12.004

Scott, S.K. (1998). The point of P-centres. *Psychological Research,* 61(1), 4–11.

Scott, S.K., McGettigan, C., & Eisner, F. (2009). A little more conversation, a little less action—candidate roles for the motor cortex in speech perception. *Nature Reviews Neuroscience,* 10(4), 295–302. DOI: 10.1038/nrn2603

Skipper, J.I., van Wassenhove, V., Nusbaum, H.C., & Small, S.L. (2007). Hearing lips and seeing voices: how cortical areas supporting speech production mediate audiovisual speech perception. *Cerebral Cortex,* 17(10), 2387–2399. DOI: 10.1093/cercor/bhl147

Ten Oevers, 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(2016), 1–9. DOI: 10.1016/j.neuroimage.2016.07.011

Ten Oevers, S., & Sack, A.T. (2015). Oscillatory phase shapes syllable perception. *Proceedings of the National Academy of Sciences of the United States of America,* 112(52), 15833–7. DOI: 10.1073/pnas.1517519112

Treiman, R., & Danis, C. (1988). Syllabification of intervocalic consonants. *Journal of Memory and Language,* 27(1), 87–104.

Wagner, P., & Dellwo, V. (2004). Introducing YARD (Yet Another Rhythm Determination) and re-introducing isochrony to rhythm research. *Proceedings of Speech Prosody 2004*, 227–230.

Wilson, M., & Wilson, T.P. (2005). An oscillator model of the timing of turn-taking. *Psychonomic Bulletin & Review,* 12(6), 957–968. DOI: 10.3758/BF03206432

Wilson, S.M., Saygin, A.P., Sereno, M.I., & Iacoboni, M. (2004). Listening to speech activates motor areas involved in speech production. *Nature Neuroscience,* 7(7), 701–702. DOI: 10.1038/nn1263

Zion Golumbic, E., Cogan, G.B., Schroeder, C.E., & Poeppel, D. (2013). Visual input enhances selective speech envelope tracking in auditory cortex at a "cocktail party". *The Journal of Neuroscience,* 33(4), 1417–26. DOI: 10.1523/JNEUROSCI.3675-12.2013