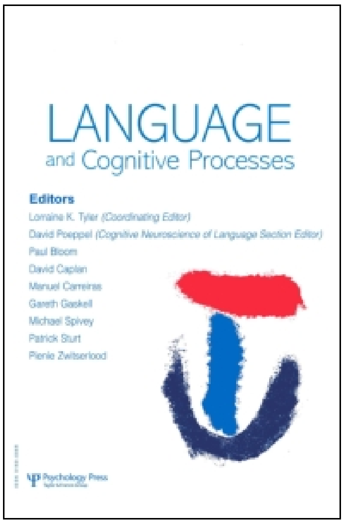

Language and Cognitive Processes
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713683153

The word segmentation process as revealed by click detection

David M. Gómez[ab]; Ricardo A. H. Bion[ac]; Jacques Mehler[a]
[a] Cognitive Neuroscience Sector, International School for Advanced Studies (SISSA/ISAS), Trieste, Italy
[b] Department of Mathematical Engineering (DIM), University of Chile, Santiago, Chile [c] Department of Psychology, Stanford University, Stanford, CA, USA

First published on: 07 June 2010

To cite this Article Gómez, David M. , Bion, Ricardo A. H. and Mehler, Jacques(2011) 'The word segmentation process as revealed by click detection', Language and Cognitive Processes, 26: 2, 212 — 223, First published on: 07 June 2010 (iFirst)
To link to this Article: DOI: 10.1080/01690965.2010.482451
URL: http://dx.doi.org/10.1080/01690965.2010.482451

LANGUAGE AND COGNITIVE PROCESSES
2011, 26 (2), 212–223

Psychology Press
Taylor & Francis Group

The word segmentation process as revealed by click detection

David M. Gómez[1,2], Ricardo A. H. Bion[1,3], and Jacques Mehler[1]

[1]Cognitive Neuroscience Sector, International School for Advanced Studies (SISSA/ISAS), Trieste, Italy
[2]Department of Mathematical Engineering (DIM), University of Chile, Santiago, Chile
[3]Department of Psychology, Stanford University, Stanford, CA, USA

The click detection paradigm was once conceived as a method to study online syntactic processing, but well-controlled empirical investigations casted many doubts on the early findings based on it. In this paper, we show that click methods can still prove valid and useful. We asked adult participants to listen to an artificial speech stream composed of statistically defined trisyllabic nonce words, while having to detect clicks superposed on the stream. The clicks were presented either within or between consecutive words. After 2 minutes of exposure to the stream, participants were slower to detect clicks located within words than clicks located between words. This result suggests that methods like click detection are sensitive to online statistical computations, opening new possibilities to obtain a richer picture of the segmentation process than what was hitherto possible.

Keywords: Speech segmentation; Artificial speech streams; Transitional probabilities; Click detection.

Correspondence should be addressed to David M. Gómez, Cognitive Neuroscience Sector, International School of Advanced Studies, Via Bonomea 265, 34136 Trieste, Italy. E-mail: dgomez@sissa.it

The authors are grateful to Gareth Gaskell, Grace Budde, and two anonymous reviewers, for their helpful comments on previous versions of this draft. This research has been supported by McDonnell Foundation Grant 21002089.

---

http://www.psypress.com/lcp DOI: 10.1080/01690965.2010.482451

Efficient behavioural methods that uncover the progress of mental processes are crucial for the development of cognitive science. Some of those methods rely on offline processing measures, whereas others rely on online measures. Because the latter supposedly provide an early and often indirect measure of the underlying processing, they might be useful for exploring cognitive processes like statistical language learning.

In natural speech, most words are not separated by silence or other such salient cues; thus, a question that has been extensively investigated in recent years is how words are segmented from the continuous speech stream. In a series of widely cited studies, Saffran and colleagues (Saffran, Aslin, & Newport, 1996; Saffran, Newport, & Aslin, 1996) reported that both infants and adults are able to extract words[1] from monotone, artificial speech streams that do not signal word boundaries with pauses or other prosodic properties. Based on these results, they argued that participants succeeded in parsing such streams because of their sensitivity to the transitional probabilities (TPs) between syllables. Later studies have shown that these statistical computations are both domain (e.g., Saffran, Johnson, Aslin, & Newport, 1999) and modality (e.g., Kirkham, Slemmer, & Johnson, 2002) general. Moreover, these computations can also be done over some (but not all) segments. Bonatti, Peña, Nespor, and Mehler's (2005, 2007) studies on vowels and consonants have shown that TPs between consonants are readily used for parsing purposes, whereas TPs between vowels are not. In addition, we now know that this capacity to profit from statistical properties of speech is not unique to humans—Hauser, Newport, and Aslin (2001) showed that cotton-top tamarins can also segment statistically structured continuous speech streams.

In spite of the attention that statistical learning has received, most of the studies have assessed word extraction using offline methods. The standard use of two alternative forced choice tests with adults and the head-turn procedure with infants has prevented researchers from investigating the time course of the segmentation process. Nevertheless, understanding the time course of this process might be as interesting for research as knowing how the extraction of words through the use of TPs happens and it could reveal constraints relevant for any statistical learning candidate model (especially computational models).

Mainstream offline methods have left some unresolved issues regarding the amount of exposure that is required before segmentation arises with different age groups. Adults are generally exposed to at least 10 minutes of the speech

[1] When referring to artificial speech streams, hereafter we call "words" the speech units that can be extracted from a speech stream due to the high TPs between adjacent syllables, and the drop in TPs between the last syllable of a unit and the first syllable of the next unit. In order to isolate statistical computations, these words are usually meaningless for the listener. For streams composed with nonspeech units, the same notion applies in a straightforward manner.

stream before being tested on segmentation, whereas infants are typically exposed to 2 minutes prior to testing. Some studies, such as Peña, Bonatti, Nespor, and Mehler (2002), have used short exposure periods with adult participants, but their material included subliminal pauses between words since their interest was to assess rule learning rather than segmentation. Endress and Bonatti (2007) proposed that rule learning is usually carried on the basis of sparse data, whereas statistical computations take more time to produce above-chance performance in offline tests. Thus, offline methods have been unable to disentangle whether adult participants require extra exposure for statistical computations, or if it is just an artifact of the testing method.

Neurophysiological research has assessed both offline (e.g., McNealy, Mazziotta, & Dapretto, 2006, fMRI with adults; Teinonen, Fellman, Näätänen, Alku, & Huotilainen, 2009, event-related potentials [ERPs] with newborns) and online[2] (e.g., Abla, Katahira, & Okanoya, 2008; Abla & Okanoya, 2009; Buiatti, Peña, & Dehaene-Lambertz, 2009; Turk-Browne, Scholl, Chun, & Johnson, 2009) statistical learning, shedding light on the neural processes underlying segmentation. Online studies coincide in that the neural mechanisms underlying statistical computations and segmentation yield evidence of learning several minutes before performance in offline tests can be assessed: Abla et al. (2008) and Abla and Okanoya (2009) measured ERPs while participants listened to a stream of pure tones or observed a stream of geometrical figures, respectively. In both studies, participants who performed best in an offline test had enhanced N400 components at middle frontal and central electrodes for "words" (cohesive sequences of three consecutive units) instead of single unit onsets. Importantly, this ERP pattern was only present during the first minutes of continuous exposure to the statistical streams. In addition, Turk-Browne et al. (2009) showed that both lateral occipital and left ventral occipito-temporal cortices were more activated for statistically structured as compared to randomly structured blocks of visual stimuli. They also found several other brain regions that displayed activation preferentially for statistically structured blocks after 2 or 3 minutes of exposure. Buiatti et al. (2009) measured electroencephalogram (EEG) oscillatory power while participants listened to statistically or randomly structured speech streams as well. They observed a decrease of oscillatory power at the frequency corresponding to single syllables for structured, compared to random, speech streams. In addition, when they exposed participants to speech streams composed of trisyllabic sequences separated by subliminal pauses, they found greater oscillatory power at the frequency corresponding to trisyllabic

[2] Several researchers use the term "online" in different ways. For our purposes, it will apply to studies assessing the time course of segmentation over a single experimental session. This requires providing at the minimum two measures indexing segmentation along that session, in order to approach (at least partially) the dynamics of the segmentation process.

sequences in streams composed of words than in random ones. These effects were observed at the end of the third minute of exposure.

In this paper, we borrow a behavioural method used in psycholinguistics to investigate online sentence processing—specifically, our measure was the latency with which clicks are responded to when superposed on speech streams. We investigated whether click detection could provide an index of online statistical computations and speech segmentation.

Some decades ago, psycholinguists devised a click location paradigm in an attempt to investigate sentence processing. In this method, participants are asked to listen to a sentence and then report the location of a click superposed to it. Fodor, Bever, and Garrett used this method extensively (e.g., Fodor & Bever, 1965; Fodor, Bever, & Garrett, 1974; Garrett, Bever, & Fodor, 1966) to study syntactic processing. Concurrently, other methods like phoneme detection, word detection, click detection, and tone detection were also employed. A common underlying assumption was that reaction times (RTs) to detect the corresponding targets reflect the complexity of sentence processing in real time. Some of the early studies using these methods showed, for instance, that RTs to phonemes are longer in structurally complex sentences as opposed to structurally simple ones (Foss & Lynch, 1969) and that RTs to clicks located in major syntactic boundaries are shorter than RTs to clicks not in a boundary (Holmes & Forster, 1970). However, findings obtained from different detection tasks were not always consistent (for a review of the early results of detection methods, see Cutler & Norris, 1979). For example, RTs to phonemes in stressed syllables are shorter than the ones to phonemes in unstressed syllables, whereas the reverse effect holds for RTs to clicks (Bond, 1972; Cutler & Foss, 1977). Cutler, Kearns, Norris, and Scott (1993) found that monolingual English and French speakers showed the same pattern of responses when detecting clicks within English sentences (only intelligible to the English speakers). Based on these results, they argued that click detection is primarily sensitive to acoustic properties of the linguistic material rather than to more abstract properties, such as syntax or semantics.

Nonetheless, some studies (Cohen & Mehler, 1996; Frauenfelder, Segui, & Mehler, 1980) obtained results supporting that the main effects observed in detection studies can be attributed to sentence comprehension. In both studies, the key feature affecting click RTs was the reversibility of object and subject functions in relative clauses: compare, for instance, the French sentences "Le garçon qui vit la fille" ["The boy who saw the girl"] and "Le garçon que vit la fille" ["The boy whom the girl saw"]. This is an example taken from Cohen and Mehler (1996), where by changing only one phoneme (/i/ to /ə/, and vice versa) they reversed object and subject. Frauenfelder et al. (1980) showed that reversibility of a relative clause induces different RTs to phonemes located after the clause boundary for subject and object relatives. Cohen and Mehler (1996), using click detection, replicated and extended these results: the RT difference is specific to transposed object relatives (instead of normal object relatives), and to relative sentences (rather than active or passive ones).

Click detection has been adapted also to study the role of distributional and rhythmic cues for word segmentation in 8-month-old infants. Morgan (1994; see also Morgan & Saffran, 1995) trained babies to turn their heads to a location above a loudspeaker (where a puppet would appear) each time a buzz occurred during a stream of words. In testing, babies oriented faster to the puppet location after buzzes external to cohesive units defined by the aforementioned cues.

In this study, we explore the click detection procedure as a behavioural alternative for online assessment of the segmentation process. We hypothesise that RTs to detect clicks located between consecutive words and within words will become different as statistics are computed over the speech stream. The tight acoustic and prosodic control given by speech synthesis allows minimising the low-level cues that interfered with previously obtained results, evaluating at the same time the validity of click detection paradigms in the context of artificial grammar learning.

METHODS

Participants

Twenty-eight adults participated in this experiment (11 men, aged 23 ± 3 years). They were paid for their participation, reported no auditory or language-related problems, and were naive with respect to the aims of this study. All participants were native speakers of Italian, recruited from the city of Trieste.

Stimuli

An artificial speech stream was generated using the MBROLA speech synthesiser (Dutoit, 1997) and the Italian female diphone database it4, with a sampling frequency of 16 KHz. The initial and final 5 s were ramped to avoid extra cues to segmentation. The stream contained the pseudowords: *pabuda*, *gifoto*, *venola*, and *minaro*. Each syllable lasted 240 ms, with no pauses between consecutive pseudowords. Word order was randomised such that no word was repeated twice in a row. Using the software Praat (Boersma, 2001), a set of clicks were inserted into the speech stream by modifying the audio waveform, clipping five consecutive samples of it.[3] As can be seen in Figure 1, each click

[3] By "clipping," we mean manually setting the audio waveform to its maximum possible value at a given sample. Because our clicks were produced by clipping only five consecutive samples, the speech streams were altered for less than 0.5 ms for each click. Click positions were randomly generated for the stream, not for each participant.

Figure 1. Example waveform of one of the artificial speech streams. The arrows show a click located between words (*pabuda*!*gifoto*) and a click located within a word (*pa*!*buda*).

could occur either between two words (e.g., *pabuda*!*gifoto*, where the exclamation mark indicates the location of the click) or within a word, immediately after the first syllable (e.g., *pa*!*buda*). The stream was 4 minutes long and contained 64 randomly spaced clicks, with an average interval of approximately 3.81 s (standard deviation [SD] = 1.43, range 1.68–11.28) between consecutive clicks. In order to control for acoustic properties of the selected word set, we generated another speech stream containing the words *dagifo*, *nolami*, *narove*, and *topabu*. Furthermore, to control for possible timing cues, click positions in this new stream were not the same as in the first one, although they were similarly distributed (average interval of 3.90 s between consecutive clicks, SD = 1.45, range 1.68–11.28). Notice that each one of the eight selected words actually occur in both streams. For instance, the word *gifoto* of the first stream appears also in the second stream when *dagifo* is followed by *topabu*. Half of the participants were exposed to each stream.

Procedure

Participants were tested individually in a sound-attenuated room, and listened to the material through headphones. They were instructed both to attend the speech stream (a "discourse in an alien language") trying to extract the words it contained, and to press a key as fast as possible each time they heard a click. One RT for each click was obtained. Measurements were obtained using PsyScope X Build 45 (http://psy.ck.sissa.it/).

Analysis

Nonresponded clicks and RTs longer than 1,000 ms or shorter than 100 ms were excluded from any analysis. RT data were then logarithmically rescaled for statistical analysis (descriptive statistics, ANOVAs, and paired *t*-tests), rejecting for each participant all the rescaled RTs deviating more than three SDs from the average (leading to the rejection of less than 2% of the data).

All graphics show the values converted back to normal scale, and the depicted 95% confidence intervals were adjusted, for a better visualisation of the within-subject comparisons when required, using the method proposed by Cousineau (2005) corrected by Morey (2008). Significance of post-hoc analyses was corrected in accordance with the Holm–Bonferroni method (Holm, 1979).

All analyses were carried out using MatLab 2008b (Mathworks, Inc.) and SPSS 11 (SPSS, Inc.).

RESULTS

Different reaction times (RTs) for both click locations

Figure 2(a) depicts participants' average RTs for both speech streams. A repeated measures ANOVA with between-subjects factor Stream (1 or 2) and within-subjects factor Click Location (between words or within words)

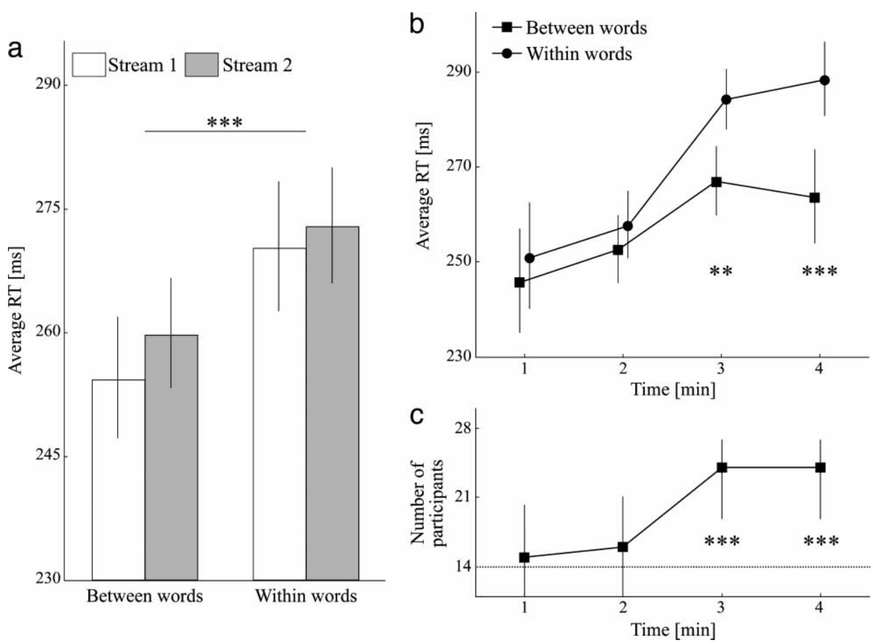

Figure 2. (a) Average RTs to clicks, segregated by Stream (1 or 2) and Click Location (between or within words). (b) Average RTs for both click locations, pooling separately RTs for each minute. (c) Number of participants for whom the average RT to clicks within words is longer than to clicks between words. The dotted line represents the chance level of 50% (14 out of 28). Vertical lines represent 95% confidence intervals, adjusted for within-subject comparisons when necessary (see Section "Methods"). $^{**}p < .01$, $^{***}p < .001$.

reveals a significant effect of Click Location, $F(1, 26) = 37.59$, $MSE = 0.04$, $p < .001$, and no significant effect of either Stream, $F(1, 26) < 1$, $MSE < 0.01$, or interaction, $F(1, 26) < 1$, $MSE < 0.01$. Thus from now on, only the pooled analyses will be presented.

Time evolution of reaction times (RTs)

To further analyse the time course of this RT difference, in Figure 2(b) we present average RTs computed minute-by-minute. A two-way repeated measures ANOVA with factors Click Location (between words or within words) and Minute (1, 2, 3, or 4) shows again a main effect of Click Location, $F(1, 27) = 26.57$, $MSE = 0.13$, $p < .001$, with RTs to clicks located within words being longer than RTs to clicks located between words. We also observe a main effect of Minute, $F(3, 81) = 16.99$, $MSE = 0.16$, $p < .001$, and a significant interaction, $F(3, 81) = 4.04$, $MSE = 0.02$, $p < .01$. Put together, these effects reveal that RTs to clicks located between words do not differ significantly across time (a one-way repeated measures ANOVA considering only the RTs for clicks located between words shows no significant effect of Minute, $F(3, 81) = 1.55$, $MSE = 0.04$), whereas RTs to clicks located within words are longer in the second half of the exposure ($t(27) = 7.92$, $p < .001$, paired t-test for average RTs in minutes 1 and 2 versus 3 and 4).

Regarding the RT differences between both click locations, a post-hoc analysis shows that these are nonsignificant during the first half of the exposure and significant in the third and fourth minutes (paired t-tests for minute 1: $t(27) = 0.99$, *ns*; minute 2: $t(27) = 1.43$, *ns*; minute 3: $t(27) = 3.88$, $p < .01$; minute 4: $t(27) = 5.09$, $p < .001$). Since no other information apart from the TPs between syllables could cue the different click locations across both streams, this result suggests that the difference in RTs is due to the discovery and perception of the statistical words present in each stream through the segmentation process.

A similar pattern appears if we consider the number of participants for which the average RT associated to clicks located within words is longer than the associated to clicks located between words. As can be seen in Figure 2(c), in minutes 3 and 4, 24 out of 28 participants are faster responding to clicks located between words than to clicks located within words ($p < .001$, Binomial tests), in contrast with 15 out of 28 for minute 1 and 16 out of 28 for minute 2.

DISCUSSION

Our results provide evidence that a click detection task can uncover the online process of word segmentation from a continuous speech stream. In particular, our study shows that throughout the experiment a significant RT difference emerges between the clicks located at the boundaries of words and the clicks located in the interior of words. Whereas RTs do not differ in the first 2 minutes of exposure, in minutes 3 and 4 participants become slower to detect clicks located within words. Congruently with this result, we observe that the group of participants who respond faster to between-words clicks than to their within-words counterparts significantly outnumbers the group showing the opposite trend (24 out of 28, starting from the third minute).

The results found here, namely, asymmetries between edge and interior positions have also been found by Hunt and Aslin (2001), where participants learned statistical information while performing a serial RT task.

The pattern of results likely reflects the statistical learning process because we used highly controlled acoustic material and two speech streams to counterbalance the target words and part-words. Our results suggest that during exposure to the auditory material, the emergence of word candidates is the main factor determining different RTs to both types of clicks. The fact that adults respond more slowly to clicks located within words can be interpreted as a tendency of participants to expect clicks to occur at the edges of word candidates. The extraction of the TPs present in the stream (0.33 across word boundaries versus 1.0 within words) might lead participants to have stronger expectations of the following syllable within words, making them less likely to expect extraneous elements such as a click. On the contrary, when the clicks occur between words, weaker expectations are present and participants might be not as surprised. It is important to notice that clicks located between units keep this status throughout the whole stream: at the beginning they are located between syllables, and as participants process statistical information they will be positioned between words. In contrast, clicks located within words are located between syllabic units at the beginning of the exposure, but the extraction of words on the basis of statistics leads them to the interior of word units. We conjecture that processing of the syllables that become integrated as a putative word resists interruption, resulting in longer RTs for clicks within words rather than shorter RTs for clicks between words. The overall trend for RTs to become longer over time might be an indication of fatigue or overload, as participants attended both to the speech material and the clicks.

Longer RTs to clicks located within words also evokes the early findings of click location studies, which showed that clicks positioned in the middle of a syntactic and/or prosodic unit are perceptually moved towards the closest unit boundary (e.g., Fodor et al., 1974). Moreover, such a preference for edges of linguistic units is reminiscent of the work on rule generalisation by Endress, Scholl, and Mehler (2005). They showed that after a familiarisation period, word structures involving adjacent repetitions were correctly generalised only when repetitions occurred in word edges, and they argued that spontaneous rule generalisation happened in this case because of a privileged perceptual saliency of word edge positions. However, it is an open issue for future research to determine whether word edges induced by statistical learning share a similar perceptual status as the edges that are physically present in the auditory stimuli.

Will the results between our method and the offline ones correlate? This is not necessarily the case: Toro, Sinnett, and Soto-Faraco (2005) and Turk-Browne, Jungé, and Scholl (2005) have shown that divided attention due to concurrent tasks can lead to poor performance in offline tests, although they lacked tools to determine if this interference affected the statistical process itself or the conscious recall of segmented sequences. Nonetheless, Saffran, Newport, Aslin, Tunick, and Barrueco (1997) reported that both children and adults succeed in offline tests even when incidentally exposed to the speech stream, during a simple concurrent task. Click detection can be a rather demanding concurrent task, due to the requirement of giving speeded responses to a stimulus that shares the same modality as speech. This leads us to hypothesise that a click detection task might interfere with offline measures of performance, but not with statistical computations themselves (as our findings show).

Our data provide the earliest evidence so far of segmentation with adult participants using behavioural measures. Our pattern of results is comparable to the ones obtained with a variety of brain imaging techniques (Abla et al., 2008; Abla & Okanoya, 2009; Buiatti et al., 2009; Turk-Browne et al., 2009), showing that behavioural methods are also able to tap the initial stages of statistical learning. Moreover, the exposure required by our participants is close to the one used in experiments with young infants for similar types of speech streams (usually 2 or 3 minutes). This fact shows that the testing method plays a central role in our ability to assess word segmentation, suggesting also that adults' abilities to detect statistical structure are not less sharp than infants'.

To the best of our knowledge, our study is the first to use a click detection task with synthesised speech stimuli. This manipulation provides a strong control of acoustics so that RTs to clicks can focus on online processing. Our results support click detection as a method sensitive also to underlying structure (statistical structure in our case) rather than just to acoustics (cf. Cutler et al., 1993). We believe that click detection is a simple behavioural approach to study artificial language learning and will be useful for studying, for instance, effects of phonological stress or real, known words in segmenting continuous speech, or the effect of sudden changes in the statistical structure of the artificial grammar.

Manuscript received 3 August 2009
Revised manuscript received 29 March 2010
First published online 7 June 2010

REFERENCES

Abla, D., Katahira, K., & Okanoya, K. (2008). On-line assessment of statistical learning by event-related potentials. *Journal of Cognitive Neuroscience*, *20*(6), 952–964.

Abla, D., & Okanoya, K. (2009). Visual statistical learning of shape sequences: An ERP study. *Neuroscience Research*, *64*(2), 185–190.

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International*, *5*(9/10), 341–345.

Bonatti, L., Peña, M., Nespor, M., & Mehler, J. (2005). Linguistic constraints on statistical computations: The role of consonants and vowels in continuous speech processing. *Psychological Science*, *16*(6), 451–459.

Bonatti, L., Peña, M., Nespor, M., & Mehler, J. (2007). On consonants, vowels, chickens, and eggs. *Psychological Science*, *18*(10), 924–925.

Bond, Z. S. (1972). Phonological units in sentence perception. *Phonetica*, *25*, 129–139.

Buiatti, M., Peña, M., & Dehaene-Lambertz, G. (2009). Investigating the neural correlates of continuous speech computation with frequency-tagged neuroelectric responses. *NeuroImage*, *44*(2), 509–519.

Cohen, L., & Mehler, J. (1996). Click monitoring revisited: An on-line study of sentence comprehension. *Memory & Cognition*, *24*(1), 94–102.

Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. *Tutorial in Quantitative Methods for Psychology*, *1*(1), 42–45.

Cutler, A., & Foss, D. J. (1977). On the role of sentence stress in sentence processing. *Language and Speech*, *20*, 1–10.

Cutler, A., Kearns, R., Norris, D., & Scott, D. R. (1993). Problems with click detection: Insights from cross-linguistic comparisons. *Speech Communication*, *13*, 401–410.

Cutler, A., & Norris, D. (1979). Monitoring sentence comprehension. In W. E. Cooper & E. C. T. Walker (Eds.), *Sentence processing* (pp. 113–134). Hillsdale, NJ: Lawrence Erlbaum.

Dutoit, T. (1997). *An introduction to text-to-speech synthesis.* Dordrecht, Netherlands: Kluwer Academic.

Endress, A. D., & Bonatti, L. (2007). Rapid learning of syllable classes from a perceptually continuous speech stream. *Cognition*, *105*(2), 247–299.

Endress, A. D., Scholl, B. J., & Mehler, J. (2005). The role of salience in the extraction of algebraic rules. *Journal of Experimental Psychology: General*, *134*(3), 406–419.

Fodor, J. A., & Bever, T. G. (1965). The psychological reality of linguistic segments. *Journal of Verbal Learning and Verbal Behavior*, *4*, 414–420.

Fodor, J. A., Bever, T. G., & Garrett, M. F. (1974). *The psychology of language.* New York: McGraw-Hill.

Foss, D. J., & Lynch Jr., R. H. (1969). Decision processes during sentence comprehension: Effects of surface structure on decision times. *Perception & Psychophysics*, *5*, 145–148.

Frauenfelder, U., Segui, J., & Mehler, J. (1980). Monitoring around the relative clause. *Journal of Verbal Learning and Verbal Behavior*, *19*, 328–337.

Garrett, M. F., Bever, T. G., & Fodor, J. A. (1966). The active use of grammar in speech perception. *Perception & Psychophysics*, *1*, 30–32.

Hauser, M. D., Newport, E. L., & Aslin, R. N. (2001). Segmentation of the speech stream in a nonhuman primate: Statistical learning in cotton-top tamarins. *Cognition*, *78*(3), B53–B64.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, *6*(2), 65–70.

Holmes, V. M., & Forster, K. I. (1970). Detection of extraneous signals during sentence recognition. *Perception & Psychophysics*, *7*, 297–301.

Hunt, R. H., & Aslin, R. N. (2001). Statistical learning in a serial reaction time task: Access to separable statistical cues by individual learners. *Journal of Experimental Psychology: General*, *130*(4), 658–680.

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition*, *83*(2), B35–B42.

McNealy, N. Z., Mazziotta, J. C., & Dapretto, M. (2006). Cracking the language code: Neural mechanisms underlying speech parsing. *The Journal of Neuroscience*, *26*(29), 7629–7639.

Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Tutorial in Quantitative Methods for Psychology*, *4*(2), 61–64.

Morgan, J. L. (1994). Converging measures of speech segmentation in preverbal infants. *Infant Behavior and Development*, *17*, 389–403.

Morgan, J. L., & Saffran, J. R. (1995). Emerging integration of sequential and suprasegmental information in preverbal speech segmentation. *Child Development*, *66*(4), 911–936.

Peña, M., Bonatti, L., Nespor, M., & Mehler, J. (2002). Signal-driven computations in speech processing. *Science*, *298*, 604–607.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*, 1926–1928.

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, *70*(1), 27–52.

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language*, *35*(4), 606–621.

Saffran, J. R., Newport, E. L., Aslin, R. N., Tunick, R. A., & Barrueco, S. (1997). Incidental language learning: Listening (and learning) out of the corner of your ear. *Psychological Science*, *8*(2), 101–105.

Teinonen, T., Fellman, V., Näätänen, R., Alku, P., & Huotilainen, M. (2009). Statistical learning in neonates revealed by event-related brain potentials. *BMC Neuroscience*, 10:21. doi:10.1186/1471-2202-10-21

Toro, J. M., Sinnett, S., & Soto-Faraco, S. (2005). Speech segmentation by statistical learning depends on attention. *Cognition*, *97*(2), B25–B34.

Turk-Browne, N. B., Jungé, J. A., & Scholl, B. J. (2005). The automaticity of visual statistical learning. *Journal of Experimental Psychology: General*, *134*(4), 552–564.

Turk-Browne, N. B., Scholl, B. J., Chun, M. M., & Johnson, M. K. (2009). Neural evidence of statistical learning: Efficient detection of visual regularities without awareness. *Journal of Cognitive Neuroscience*, *21*(10), 1934–1945.