Temporal Processing for Syntax Acquisition: A simulation study

Jean-Marc Blanc (blanc@isc.cnrs.fr)
Christelle Dodane (dodane@isc.cnrs.fr)
Peter Ford Dominey (dominey@isc.cnrs.fr)
Institut des Sciences Cognitives
UMR 5015 CNRS-Université Claude Bernard Lyon 1
67, boulevard Pinel 69675 BRON cedex

Abstract

Early perceptual processing capabilities are likely to contribute to the categorization of lexical vs. grammatical words by newborns. This lexical categorization could be performed by detecting differences in the prosodic structure of these word categories. Here we demonstrate that a Temporal Recurrent Network (TRN) that allows realistic treatment of the dynamic temporal aspect of prosody performs this lexical categorization task on French and English. We then examine the functional relation between this capability, and non-linguistic temporal discrimination. We reduce sensitivity to temporal structure in the TRN by increasing the network time constants. This yields (1) a reduction in performance in the lexical categorization task, and (2) a deficit in the processing of brief auditory events similar to that observed for children with SLI. While our principle result is that the TRN can perform lexical categorization based on prosodic structure, it is of interest that the impaired TRN suggests a functional link between impaired temporal processing, and impaired lexical categorization in SLI.

Introduction

One of the most challenging questions in child language acquisition is how children learn syntax. Acquiring language involves classifying lexical items into syntactic categories. Four main sources of information in linguistic input have been proposed as potentially useful in this categorization: Distributional Information (Redington, Chater & Finch 1998), Semantic Bootstrapping, Phonological Constraints (Kelly 1995) and prosodic information (Shi, Werker & Morgan 1998). With respect to prosodic information coded in the fundamental frequency (F0), and lexical categorization, we note that in French, content words within a prosodic group (i.e. not at a group boundary) are marked with F0 peaks, and the F0 of content words in English are marked with a principal accent (Hirst & Di Cristo 1998).

The current research tests the hypothesis that these aspects of the temporal structure of prosody can be processed by a temporal recurrent network (TRN), thus providing insight into lexical categorization and language acquisition.

Prosodic Foundations for Syntax

Jusczyk et al. (1996) proposed that infants are able to make immediate use of their sensitivity to prosodic markers as a means for organizing the input. We will now examine several investigations that have been realized to confirm this hypothesis.

Sensitivy to prosodic structure

Adults: Bagou et al. (2002) evaluated the relative contribution of two prosodic cues, lengthening and F0 contour, in the processes of speech segmentation and storage of new words. Their results showed that prosodic information facilitates the acquisition of a new mini-language. Kelly (1995) demonstrated that American subjects were able to exploit phonological cues to identify unknown words as verbs or nouns.

Children: Though children frequently fail to produce function morphemes in their earliest utterances, Gerken and McIntosh (1993) have suggested that children by the age of 2 years have a representation of some specific function morphemes, and the context in which they appear. We will now consider the ability of newborns to exploit prosodic cues for the early basis of syntax.

Newborns: French newborns can discriminate between English and Japanese sentences (different rhythm classes), but not between Dutch and English sentences (same rhythm class). This discrimination was possible despite the speech being low-pass filtered (at 400 Hz), highlighting the role of prosody (Nazzi, Bertoncini & Mehler, 1998). Moreover, newborn infants are able to use probabilistic combinations of acoustic and phonological information to perceptually separate English word tokens into function/content categories (Shi et al. 1999). At a later stage, such an ability could potentially help bootstrap infants into acquisition of grammar by allowing them to detect and represent classes of words on the basis of perceptible surface cues. However, the specific cues used by newborns to make this distinction have not been precisely determined.

Simulation Shi et al. (1998) investigated if various “presyntactic cues” (such as number of syllables, presence of a complex syllable nucleus, presence of syllable coda, and syllable duration, to name only a few phonologically relevant cues) are sufficient to guide the assignment of words to rudimentary grammatical categories. Their investigation of English, Mandarin Chinese and Turkish shows that “sets of distributional, phonological, and acoustic cues distinguishing lexical and functional items are available in infant-directed speech across such typologically distinct languages as Mandarin and Turkish” (Shi et al. 1998). Thus grammatical words tended to be acoustically and/or phonologically minimized in comparison to lexical words.

Durieux and Gillis (2000) proposed an artificial learning system for lexical categorisation with English, based on phonological and prosodic information. However in these two simulation studies, a number of specific cues were extracted from the speech by a human expert. Here, we investigate whether a neuro-realistic system can automatically extract cues from the contour of the fundamental frequency itself.

Material & Methods

Corpora

LSCP: This corpus contains 54 French sentences read by a single native speaker. The segmentation provided groups of adjacent content and function words (~ 200 for each category; Ramus et al. 1999). The corpus served in part for Experiments 1 and 3.

MULTEXT Experiment 2 used French and English speech from the MULTEXT multilingual corpus developed for the study of prosody (Campione & Veronis, 1998). Stories were read by 20 different speakers (5 males and 5 females per language) which lead to a total of 8236 words for English, and 6945 words for French.

Fundamental Frequency (F0)

Extraction Fundamental frequency (Raw data) was obtained from the speech signal autocorrelation (PRAAT software; Boersma 1993). This description of F0 was provided as input to the TRN.

Processing To obtain an acceptable perceptual representation of intonation based on raw values of F0, we apply the MOMEL quadratic spline smoothing algorithm (Hirst & Espesser (1993), yielding a smooth continuous curve reflecting the intonation contour with which the utterance is spoken.

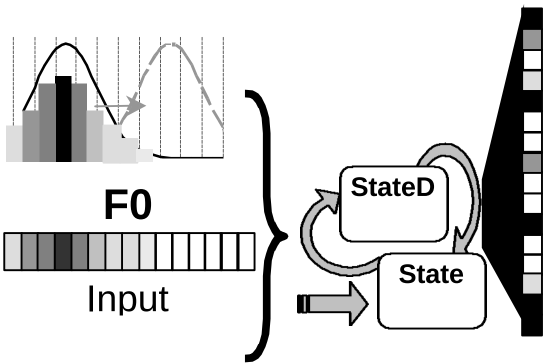
Temporal Recurrent Network (TRN)

Architecture: The TRN is based on the dynamic coding properties of recurrently connected leaky integrator neurons with sigmoid output functions that are organized in three layers: the Input, consisting of sixty units, that transmits prosodic information to the State layer, which projects to the StateD layer (See Figure 1). Recurrently connected State and StateD layers are each made up of 25 units, and StateD neurons include several different time constants to allow sensitivity to different time scales.

Principles: The connections between the different layers are randomly initialized (-0.5; +0.5) at the outset and remain fixed. Weights are then selected from a population of networks based on performance. Learning consists of evaluating the representation of input sequences in the activities of State and StateD after a sequence has been presented.

Treatment of prosody: The dynamic system that results from this recurrent network has been demonstrated to be sensitive to the temporal and serial order of input sequences, and was shown to discriminate between languages of different rhythm classes based on their prosodic structure (Dominey & Ramus 2000).

Figure 1: Temporal Recurrent Network Architecture



Representation of inputs F0 was coded in Hz at 5ms sampling intervals as a population of activity with a Gaussian distribution, where the unit with the highest activation encodes the mean frequency.

Exp. 1: Development of Lexical Identification

The purpose of Experiment 1 was the identification, based on peaks in F0, of two lexical categories: **Content words** that have a meaning-related component such as nouns, verbs, adjectives, and adverbs, and **Function words** that are primarily structural, such as articles, prepositions, and auxiliaries.

Exp 1a: Linguistics Investigations

The location of F0 peaks is based on the detection of a change in the sign of the difference between two adjacent values of F0. Thus local minima and maxima (F0 peaks) are detected. To eliminate irrelevant peaks due to deviations in the raw F0 data, the MOMEL smoothing algorithm was applied. We observed that 92% of F0 peaks and 61% of F0 valleys occurred in content words, with the opposite distribution for function words. Applying the rule that the presence of a F0 peak indicates that a word belongs to Content category, whereas the absence of this cue revealed a Function word yields the results in Table 1, where chance is 50,6%.

Table 1: French lexical categorization with LSCP

Corpus	Method	F0	% correct
LSCP French	Random		50,6
	F0 peaks	Raw data	62,1
	F0 peaks	MOMEL	82,8
	TRN	Raw data	83,6
	TRN+SOM	Raw data	84,1

Exp 1b: Temporal Neuro-inspired Simulations

The TRN was trained on half of the function (F) and content (C) words at group boundaries. After presentation of a F or C word, the vector of activity in State/StateD was sampled, and averaged into a Prototype vector, yielding prototype vectors for F and C respectively. Validation consisted of presenting the second half of the corpus, and identifying words based on their distance from the prototypes. We also conducted experiment with a Self-Organized Map to classify these vectors in a unsupervised way. Table 1 presents performance for the best of a population of 50 networks on the validation, for supervised (TRN) and unsupervised learning (TRN+SOM).

Exp 2: Application to larger Corpora

Experiment 2 applies the Peak and TRN methods of identification to a larger corpus (MULTEXT) with ten different speakers. The speech signal is divided into words rather than groups of words belonging to the same lexical category and the speech is closer to unconstrained speech than that in the LSCP corpus.

French

Table 2 demonstrates that the TRN performs the Content/Function distinction within the same range as the explicit detection of F0 peaks, superior to chance. The LSCP and MULTEXT French corpora yielded different performance, and their syntactic content differs notably. The LSCP corpus contains only sentences between 15 and 21 syllables, whereas the MULTEXT corpus is based on short passages including syntactic structure specific to spontaneous speech. Nonetheless, despite these differences, lexical categorization remains possible.

English

We subsequently tested whether F0 peaks were also characteristic of content words in English. Table 2 indicates that this intonation contour could be used to obtain the same lexical identification in English. We conclude that, the TRN can perform lexical categorization at a level comparable to human infants (Shi et al. 1999). However, as pointed out by Shi et al. 1998, the contribution of a given cue to lexical categorization may vary across languages. Thus the presence of an F0 peak may have less impact in English than in French.

Table 2: Lexical categorization with MULTEXT

Language	Method	F0	% correct
French	Random		52
	Peaks	MOMEL	73,1
	TRN	Raw data	70,3
English	Random		54
	Peaks	MOMEL	64,5
	TRN	Raw data	62,8

Impaired Prosodic Foundations for Syntax

In contrast to normally developing children, there exists a subgroup of children characterized by a significant limitation in reading and/or language development and ability without the presence of an overt underlying condition such as low overall IQ or impaired hearing. This condition is often referred to as Specific Language Impairment (SLI). Several investigators have proposed that individuals with SLI have special problems with the acquisition of functional categories (Eyer & Leonard, 1995; Leonard 1995; Bishop 1979). Van der Lely (1996) suggested that children with SLI had greater difficulty assigning roles such as agent and theme or theme and goal on the basis of syntactic structure alone.

The origin of the processing deficits in language impaired populations is still a question of debate. To summarize, it could be due to basic sensory processing deficits (Tallal et al. 1993) or to higher-level cognitive processes. The first hypothesis postulated that abnormalities in the neurophysiological encoding of acoustic differences in speech are responsible for this deficit. This impairment of auditory temporal processing was tested in several auditory experiments employing non-verbal stimuli. When stimuli are either brief or rapid, children with SLI have difficulty in discriminating them, although they have no difficulty in differentiating the same stimuli when they are lengthened or presented at slower rate. At least a subgroup of children with SLI exhibits a deficit for processing rapid auditory events at the same time that they have trouble with the manipulation of function words and phonemes (Tallal 1998).

These observations lead us to three important experimental considerations: Exp 1a: Prosodic cues can allow the distinction of function versus content words; Exp 1b and Exp 2: This prosodic structure can be processed by a temporal neural network model to perform this identification; Exp 3: A deficit in temporal prosodic processing could produce impairments both in lexical categorization and in auditory temporal discrimination tasks based on pure tones.

Exp. 3: Simulation of a Temporal Deficit

In Exps 1 and 2, the processing of temporal context likely plays an important role, as F0 peaks are defined within a progressive rising and falling context, implying a required sensitivity to peaks and the surrounding temporal context. In this context, the TRN is sensitive to serial and temporal structure. Thus, the objective of this experiment is to determine if perturbation of the temporal sensitivity of the network will yield the same impairments as observed in children with SLI for the manipulation of syntax (i.e. lexical categorization), and also for auditory processing of rapid events.

Impaired Lexical Categorization

A population of 50 control networks was able to perform lexical categorization with a mean of 75% correct on the LSCP corpus. To reduce sensitivity to temporal structure in the TRN we increased the network time constants to produce 50 distorted networks. None of the distorted networks were able to perform the categorization task (mean of 50%, and maximum for 53%), whereas these models could perform a sequence identification task based on 3-element sequences with elements of long duration (800 ms).

Auditory Repetition Task

The Rapid Perception Test was elaborated to test the ability of children with SLI to handle rapid auditory events. Tallal and Piercy (1973) demonstrated that normal children were able to identify the order of two 75ms tones separated by an inter-stimulus interval (ISI) as short as 8ms, while individuals with SLI required an ISI exceeding 300ms (Fig 2).

These experiments were replicated on the two populations of TRN (control and distorted), based on comparing the State/StateD vectors in a discrimination measure. Figure 2 reveals that the same disorder was observed for the distorted network which failed in the case of short ISIs. As observed in children with SLI, performance improves rapidly for longer ISI durations for the distorted networks.

Backward Masking

Wright et al. (1997) tested the hypothesis that the auditory temporal processing deficits observed in SLI children could be related to impairments in the detection of temporally interacting stimuli as in tasks of forward and backward masking. They reported that control children showed a significantly lower tone thresholds than SLI children for detecting a tone in the backward masked condition, in which the target tone immediately follows a white noise mask.

The model performs this backward masking task in which the presence or absence of a 20 ms tone (coded by activation of specific F0 units) should be detected after a 300 ms noise mask (coded by uniform activation of F0 units) (Fig 3). Wright et al. (1997) demonstrated that children with SLI required significantly higher intensity levels for the tone than normal children to perform this task. This experiment reveals the same distinction between the normal and distorted networks, as between normal and SLI children, as illustrated in Figure 3.

Discussion

Implication for a Simulated Acquisition of Language

The first two experiments demonstrated that the intonation contour could serve as basis for an identification of Function vs. Content words, that could bootstrap the acquisition of syntax. Furthermore we demonstrated that a biologically plausible recurrent network system could extract reliable information from the fundamental frequency, without human labeling of speech samples (as was required in Durieux & Gillis 2000; Shi et al. 1998). This observation has been demonstrated on two languages, English and French.

In this context it would of interest to study Child Directed Speech (CDS), that is known to including more salient information. Indeed CDS displays an exaggeration of prosody, in particular in the contour of F0. In addition to previous work, supplementary information (like a representation of spectrum) could be provided to the network to increase the performance of the lexical categorization.

Neurophysiological Implication

The TRN is a an extension of a model originally developed to simulate primate behavioral electrophysiology experiments (Dominey et al. 1995). In the current context it demonstrated how temporal processing can contribute a rudimentary basis for syntax in the form of lexical categorization. This confirmed the TRN's utility in processing prosodic and distributional regularities (Dominey & Ramus 2000).

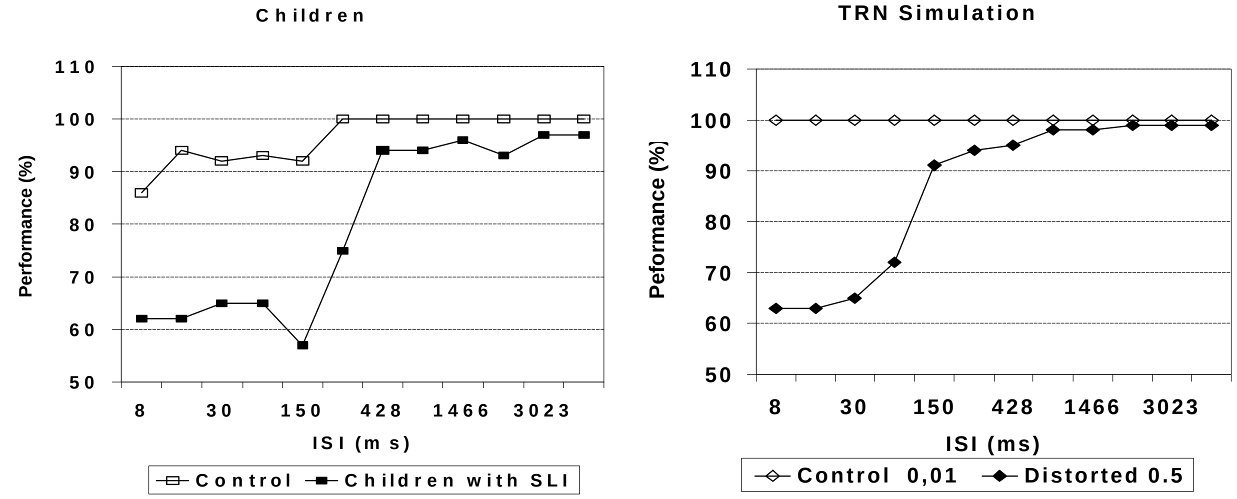


Figure 2: Comparison of children performance versus TRN on the Auditory Repetition Task (Tallal & Piercy 1973)

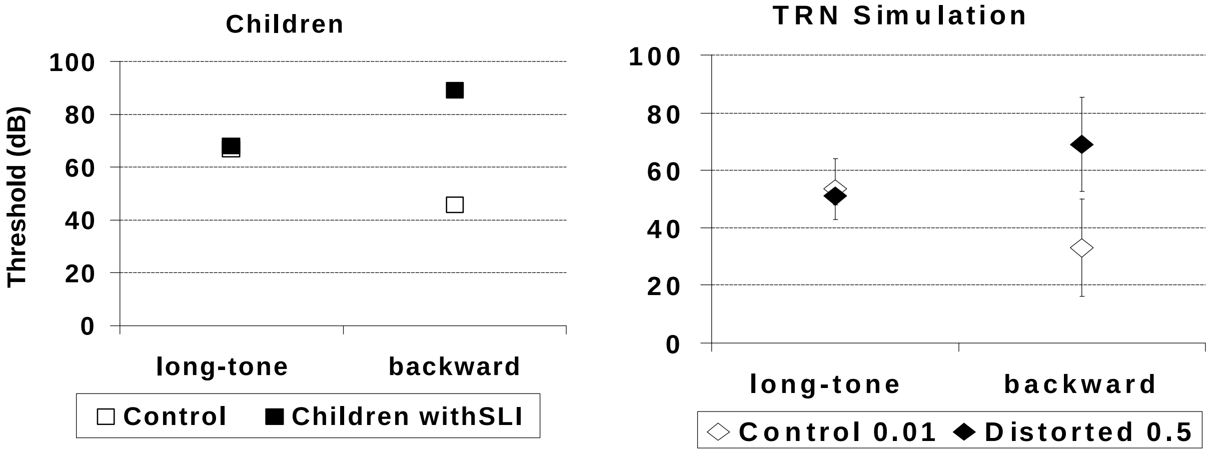


Figure 3: Tone threshold for TRN and children for the first two conditions of the task in Wright et al. 1997.

All these experiments lead to the conclusion that a fine-grained treatment of temporal structure in speech is of primary importance for initiating syntax acquisition.

Deficits in Prosodic Processing, and SLI

Several studies have reported the difficulties faced by children with SLI with manipulation of function words, together with a deficit in treatment of rapid auditory events. The time window implied in this deficient processing has been proposed to be roughly the duration of a syllable (Tallal & Piercy, 1973). In this case it will be difficult to localize F0 peaks within the word structures of sentences. In this context the TRN failed to identify Content and Function words, when its sensibility to fine temporal structure was altered. Tallal and al. (1998) indicated that this auditory deficit was reversible, indeed after training on the ART task, thresholds approached the normal tens-of-ms range. These results pose interesting simulation goals, and we are currently investigating the effects of training and discrimination threshold modifications for improving task performance in our SLI simulations.

Conclusion

Three objectives were achieved with these experiments. First, two basic categories (Function and Content Words) could be identified from the prosodic structure of speech. Second, this distinction could be performed by a Temporal Recurrent Network, that was developed in a functional neurophysiology framework (Dominey et al. 1995). Third, the TRN could be perturbed to simulate the behavior of children with SLI for detection of rapid auditory events, while at the same time its lexical categorization based on prosody is severely attenuated. This provides the basis for future studies of the contribution of temporal structure of speech to normal and impaired language acquisition.

Acknowledgments

JMB is supported by a Doctoral fellowship from the Région Rhone-Alpes, CD by a Post-Doctoral fellowship from the HFSP Organization, and PFD by the ACI Integrative and Computational Neuroscience, the OHLL and the Eurocores OMLL Projects. We thank F. Ramus, A. Christophe and E. Dupoux for insightful discussion and access to the LSCP corpus.

References

Bagou, O., Fougeron, C., & Frauenfelder, U.H. (2002). Contribution of Prosody to the Segmentation and Storage of "Words" in the Acquisition of a New Mini-Language, *Prosody 2002*, 11-13 April 2002, 159-162.

Bishop, D.V.M. (1979). Comprehension in developmental language disorder. *Developmental Medicine and Child neurology,* 21, (pp. 225-238).

Boesrma, P., Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. *Proceedings of the Institute of Phonetic Sciences* 17, pp. 97-110.

Campione, E., & Véronis, J., (1998). A multilingual prosodic database, *Proc. of ICSLP'98, Sidney*.

Dominey, P.F., Arbib M.A., & Joseph J.P. (1995). A Model of Cortico-Striatal Plasticity for Learning Oculomotor Associations and Sequences, *Journal of Cognitive Neuroscience*, 7:3, 311-33.

Dominey P.F., & Ramus F. (2000). Neural network processing of natural language: I. Sensitivity to serial, temporal and abstract structure of language in the infant. *Language and Cognitive Processes*, 15(1) 87-127.

Durieux, G., & Gillis, S. (2000). Predicting grammatical classes from phonological cues: An empirical test. In *Approaches to bootstrapping: phonological, syntactic and neurophysiological aspects of early language acquisition*, ed. B. Höhle, J. Weissenborn, pp. 189-232. Amsterdam: Benjamins

Eyer J., Leonard L., (1995). Functional categories and specific language impairment: A case study. *Language Acquisition*, 4, 177-203.

Gerken, L. A., & McIntosh, B. J. (1993). The interplay of function morphemes and prosody in early language. *Developmental Psychology*, 29, 448-457

Hirst D., & Di Cristo A., (1998). A survey of intonation systems. in Hirst & Di Cristo (eds). *Intonation Systems : A Survey of Twenty Languages*.,1-44.

Hirst, D. & Espesser, R.(1993). Automatic modelling of fundamental frequency using a quadratic spline function. *Travaux de l'Institut de Phonétique d'Aix* 15, 71-85.

Jusczyk, P.W., Kelmer Nelson, D.G. (1996). Syntactic units, prosody, and psychological reality during infancy, In JL Morgan, K Demuth (Eds) *Signal to Syntax: Bootstrapping from speech to grammar in early acquisition*, Lawrence Erlbaum, Mahwah NJ.

Kelly, M.H. (1995). The role of phonology in grammatical category assignments. In J.L. Morgan and K. Demuth (Eds.), *Signal to syntax: Bootstrapping from speech to grammar in early acquisition* (pp. 249-262). Mahwah, NJ: Lawrence Erlbaum Associates.

Van der Lely, H. K. J. (1996). Specifically language impaired and normally developing children: Verbal passive vs. adjectival passive sentence interpretation, *Lingua* 98-4, 243-272.

Leonard, L. (1995). Functional categories in the grammars of children with specific language impairment. Journal of Speech and Hearing Research, 38, 1270-1283. Leonard L. (1998). *Children with specific language impairment*. Cambridge, MA: MIT Press.

Nazzi, T., Bertoncini, J., & Mehler, J. (1998). Language discrimination by newborns: towards an understanding of the role of rhythm. *Journal of Experimental Psychology: Human Perception and Performance.*

Ramus, F., Nespor, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, 73(3), 265-292.

Redington, M., Chater, N. & Finch, S. (1998). Distributional information: A powerful cue for acquiring syntactic categories. *Cognitive Science*, 22, 425-469.

Shi R., Morgan J.L., & Allopenna P. (1998). Phonological and acoustic bases for earliest grammatical category asssignment : a cross linguistic perspective. *Journal of child language* 25 (1998), 169-201.

Shi R., Werker J.F., & Morgan J.L. (1999). Newborn infants' sensitivity to perceptual cues to lexical and grammatical words, *Cognition*, Volume 72, Issue 2, B11-B21.

Tallal P., & Piercy, M. (1973). Defects of non-verbal auditory perception in children with developmental aphasia. *Nature*. Feb 16;241(5390):468-9

Tallal, P. (1998). Children with language impairment can be accurately identified using temporal processing measures: A response to Zhang and Tomblin, Brain and Language, 65, 395-403 *Brain Lang*. 1999 Sep;69(2):222-9.

Tallal, P., Merzenich, M., Miller, S., & Jenkins, W. (1998). Language Learning Impairments: Integrating Basic Science, Technology and Remediation, *Experimental Brain Research*, v. 123, p 210-219.

Wright, B.A., Lombardino, L.J., King, W..M, Puranik, C.S., Leonard, C.M., & Merzenich, M.M. (1997). Deficits in auditory temporal and spectral resolution in language-impaired children. *Nature*. May 8; 387(6629):176-8.