(pp 295 -)

PROCESSING OF THE TEMPORAL ENVELOPE OF SPEECH

Ehud Ahissar[1] , Merav Ahissar[2]

[1]The Weizmann Institute of Science, Rehovot, Israel
[2] The Hebrew University, Jerusalem, Israel

The temporal envelope of speech contains low-frequency information, which is crucial for speech comprehension. This information is essential for identification of phonemes, syllables, words and sentences (Rosen, 1992). The temporal envelope of speech defines slow variations of the spectral energy of a spoken sentence, variations that are usually below 8 Hz (Houtgast and Steeneken, 1985). Comprehension of speech depends on the integrity of its temporal envelope between 4 and 16 Hz (Drullman et al., 1994; van der Horst et al., 1999). The mechanisms by which this information is extracted and processed are not yet known.

Speech comprehension does not depend on the exact frequency of the temporal envelope. In fact, the temporal envelope of normal speech can be compressed to 0.5 of its original duration before comprehension is significantly affected (Foulke and Sticht, 1969; Beasley et al., 1980). Thus, brain mechanisms normally responsible for speech perception can adapt to different input rates within a certain range (see Miller et al., 1984; Newman and Sawusch, 1996; Dupoux and Green, 1997). This on-line adaptation is crucial for speech perception, since speech rates vary between different speakers, and change according to the emotional state of the speaker.

The ability of listeners to follow fast speech rates varies significantly. In particular, reading disabled subjects (RDs) are usually more vulnerable to the time compression of sentences than good readers (Freeman and Beasley, 1978; Riensche and Clauser, 1982; Watson et al., 1990), although not always (McAnally et al., 1997). Deficiency of RDs in perceiving time-compressed speech might be related to their impaired processing of successive auditory signals (Tallal and Piercy, 1973; Aram et al., 1984; Shapiro et al., 1990; Bishop,

1992; Tallal et al., 1993; Farmer and Klein, 1995; Hari et al., 1999b; Helenius et al., 1999; Ahissar et al., 2000b; Amitay et al., 2002a; Banai and Ahissar, in press). Inadequacies of RDs in comprehending time-compressed speech appear to emerge at the cortical level (Welsh et al., 1982). One sign of such impaired processing is probably reduced cortical reactivity to speech signals (Renvall and Hari, 2002).

Here we review and interpret our previously reported data that demonstrates that auditory cortical activity in humans follows the temporal envelope of speech, and that stimulus-response temporal locking correlates with speech comprehension. We will consider two components of temporal locking: frequency matching and phase-locking. Our data is consistent with comprehesion of short sentences depending primarily on a-priori matching between the sentence envelope and an intrinsic frequency, and secondarily on on-line phase locking. Further analysis of our data demonstrates that the ability to comprehend short sentences depends on the existence of a cortical frequency mode that is equal to, or higher than, the highest frequency mode of the envelope. Moreover, accuracy of reading non-words also correlated with the maximal cortical frequency mode. The novel psychophysical data presented here demonstrate that RDs are less tolerant than good readers to accelerated speech.

Based on our data and those of others, we hypothesize that the brain sets time intervals for analysis of individual speech components by intrinsic oscillations that are pre-tuned to an expected speech rate, and re-tuned during continuous speech processing by locking to the temporal envelope. By analogy with the somatosensory system, we suggest that such temporal locking is achieved by thalamocortical phase-locked loops, which may involve the non-lemniscal thalamic pathway. Such loops also extract temporal information contained in the envelope signal and recode it by rate. Such rate-coded signals could be used for further, temporally-loose processing, and for closing the sensory-motor loop. Here, we present, schematically, the primary function of this hypothetical algorithm, which is to time syllable processing.

Comprehension of time-compressed speech by RDs and good readers

So far, deficiencies of RDs in comprehending time-compressed speech were demonstrated for time compressions of 0.5 or 0.6 of normal rate. At these compressions, speech comprehension is still above chance level, and, as a result, a complete psychometric curve could not be constructed. Thus shifts in psychometric curves of time-compressed speech between RDs and good readers could not be estimated. To obtain complete psychometric curves, we applied a novel time-scale compression algorithm that kept the spectral and pitch content intact across compression ratios down to 0.1 of the original rate (Ahissar et al., 2001).

Table 1. Sentences for psychophysics experiment (1, true; -1, false)

1.	Six is greater than three	1
2.	Three is greater than six	-1
3.	Two plus three is five	1
4.	Two plus six is nine	-1
5.	Three plus six is nine	1
6.	Three plus three is three	-1
7.	Six is smaller than three	-1
8.	Three is smaller than six	1
9.	Five and five are ten	1
10.	Five and nine are ten	-1

We tested 6 RDs (native English speakers, ages 20 - 45, with documented history of reading disability), and 7 good readers (5 post-doctorate fellows and students, and 2 staff members of the Keck Center at UCSF; 5 were native English speakers, and English was the second language of the other 2; ages 29 – 45) for comprehension of time-compressed speech. Ten sentences (Table 1) were compressed to 6 different rates (0.1, 0.2, 0.3, 0.4, 0.5, and 1 of their original duration). The compression rate of 1 (i.e., no compression) was presented to all RDs, and to 3 of the 7 good readers. Each compression was presented 20 times. Thus, subjects were given either 100 trials (compressions 0.1, 0.2, 0.3, 0.4, and 0.5) or 120 trials (compressions 0.1, 0.2, 0.3, 0.4, 0.5, and 1). Comprehension was quantified using a Comprehension Index (CI) = [($N_{correct}$ – $N_{incorrect}$) / N_{total}], where N equals the number of trials. CI could have values between –1 (all incorrect) and 1 (all correct), where 0 was the chance level.

Compared to good readers, RDs showed poorer comprehension at all speech rates, except for 0.1 at which both groups were at chance level, and 1 at which both groups showed similar performance (Fig. 1A). RDs were still at chance level at a compression of 0.2, whereas good readers started to show some comprehension, and RDs exhibited marginal comprehension at 0.3, whereas good readers already comprehended about 75% of the sentences (CI = 0.5). In fact, between compression levels of 0.1 and 0.5, the comprehension curve of RDs was shifted by 0.1 towards weaker compressions.

The impaired comprehension of the RDs appeared to be specific to three of the ten sentences used (sentences 2, 6 and 7) across most comprehension levels. CI averaged across all compression levels is depicted in Figure 1B.

Within-session improvement occurred in both the RD and good reader groups (Fig. 1C). Since compression levels, which determine task difficulty, were randomized for each subject, these levels distributed differently along the session for different subjects. Thus, comprehension normalized with regard to the average compression level presented to each subject is a better representation of within-session improvement (Fig. 1D). Examination of the two within-session comprehension plots suggests that the comprehension of RDs improves significantly during sessions, starting at chance level and approaching the performance of good readers in the second half of the session.
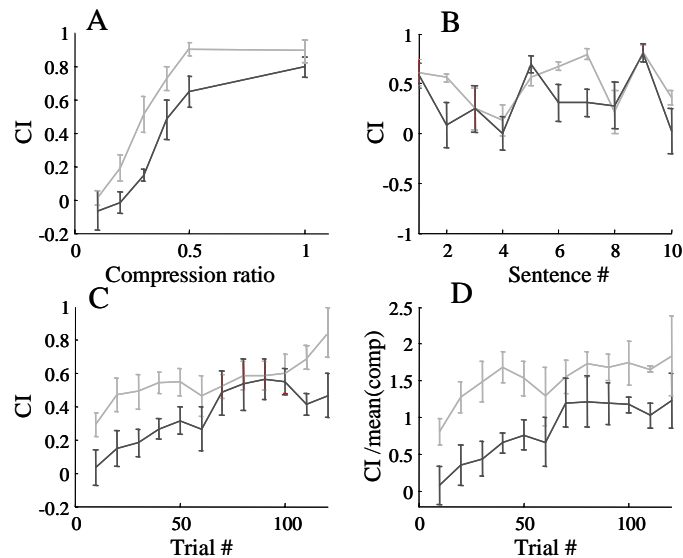
Figure 1. *Psychophysical evaluation of comprehension of time-compressed speech sentences by RDs (n=6, black) and good readers (n=7, gray). **A**. Mean comprehension as a function of compression level. **B**. Mean comprehension as a function of sentence number (Table 1). **C**. Within session improvement; mean comprehension as a function of trial number. Data were smoothed by a moving average window of 20 trials, shifted in steps of 10 trials. **D**. Within session improvement, normalized comprehension; Mean(comprehension) / mean(compression) was computed for each 20-trials window for each subject before averaging. Same smoothing as in C.*

Magnetoencephalogram (MEG) recordings

Thirteen subjects (4 good readers and 9 RDs, 7 males and 6 females, 11 native English speakers and English was the second language of the other 2, ages 25-45; one of the RDs also participated in the psychophysical experiment described above) volunteered to participate in this MEG experiment. Reading abilities were again heterogeneous, ranging from 81 to 122 in a word-reading test, and 78 to 117 in a non-word reading test, with 100 being the population average (Woodcock, 1987).

Three balanced sets of sentences were used (Table 2). Subjects were exposed to the sentences before the experiment, and the mapping between each sentence and its associated correct response (true or false) was clarified to them. During the experiment, the sentences were played at compressions of 0.2, 0.35, 0.5 and 0.75 (in some cases, compression of 1.0 instead of 0.75 was presented). Average sentence duration at compression 1.0 (i.e., without compression) was about 1 s. For each sentence, subjects responded by pressing one of three buttons corresponding to true, false, or don't know, using their left hand.

Table 2. Sentences for MEG experiment (1, true; -1, false)

Set 1

1.	Two plus six equals nine	-1
2.	Two plus three equals five	1
3.	Three plus six equals nine	1
4.	Three plus three equals five	-1

Set 2

1.	Two minus two equals none	1
2.	Two minus one equals one	1
3.	Two minus two equals one	-1
4.	Two minus one equals none	-1

Set 3

1.	Black cars can all park	1
2.	Black cars can not park	-1
3.	Black dogs can all bark	1
4.	Black dogs can not bark	-1
5.	Black cars can all bark	-1
6.	Black cars can not bark	1
7.	Black dogs can all park	-1
8.	Black dogs can not park	1
9.	Playing cards can all park	-1
10.	Playing cards can not park	1

Magnetic fields were recorded from the left hemisphere in a magnetically-shielded room, using a 37-channel biomagnetometer array with SQUID-based first-order gradiometer sensors (Magnes II; Biomagnetic Technologies, Inc.). Data acquisition epochs were 3000 ms in total duration with a 1000 ms pre-stimulus period. Data were acquired at a sampling rate of 297.6 Hz. For each subject, data were first averaged across all artifact-free trials. After which, a singular value decomposition was performed on the averaged time-domain data for the channels in the sensor array, and the first three principal components (PCs) calculated. These three PCs, which typically accounted for more than 90% of the variance within the sensor array, were used for all computations related to that subject. The basic phenomena described here could usually be observed in a single PC, usually PC1. Thus, for clarity, all examples presented here include only one PC.

In each set of sentences, envelopes of the sentences were selected such that they had similar temporal patterns (Ahissar et al., 2001). Thus, responses to different sentences could be averaged. The data presented here were obtained by averaging trials of all sentences for each compression level.

For each subject, there were 371 – 800 trials, resulting in, on the average, 93 - 200 trials per compression level. Sentences were composed in English and tested on English speakers. Since typical speech rates are similar across languages, we expect that our results regarding the processing of the temporal envelope of speech can be generalized to other languages.
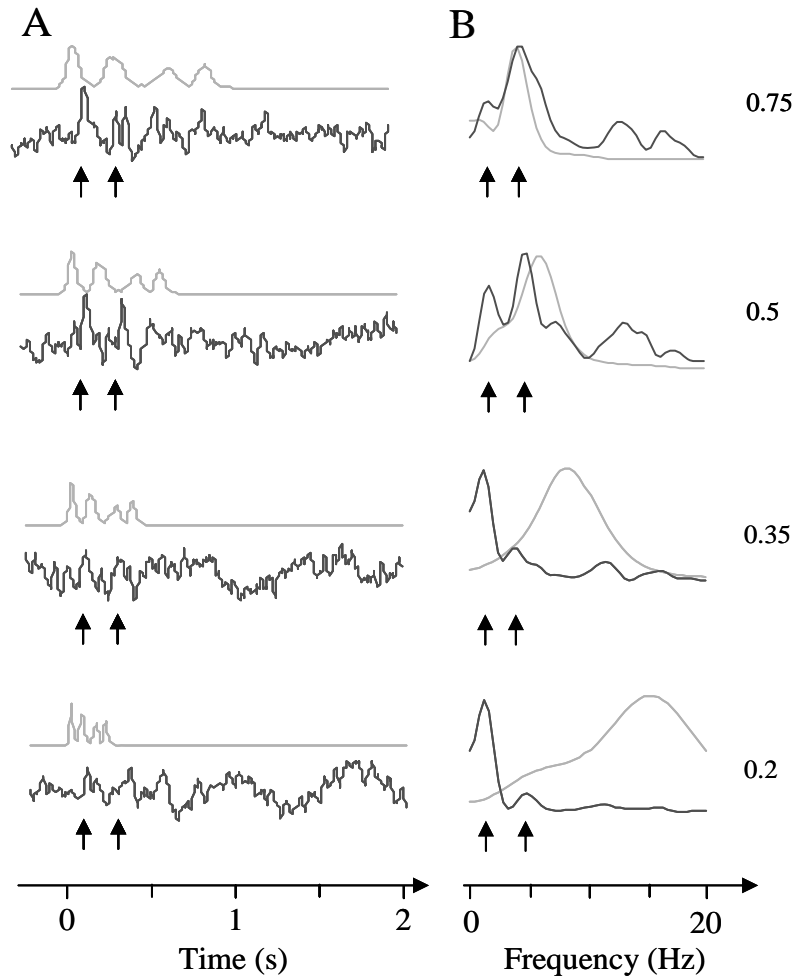
*Figure 2. Constancy of cortical frequencies. Data of PC1 of an RD subject, MS. **A**. PC1 (black) and envelope (gray) as a function of time from sentence onset for all 4 compressions. For each compression, data for all sentences were averaged. Arrows indicate the first two response peaks. **B**. Power spectra of PC1 and envelope. Arrows indicate the first two frequency modes of the response. Fast Fourier Transforms (FFTs) were computed using windows of 1 s and overlaps of 0.5 s. Power scale is 0 to max power. Compression levels indicated on the right.*

Relationship between auditory cortical activity and the temporal envelope

When responses of different trials with the same compression level were averaged, they often showed clear fluctuations between 4 to 10 Hz, with the latency of the first response peak being close to 100 ms (Fig. 2A, an RD subject). These fluctuations added up across trials, which indicates that they are time-locked to the stimulus: either each of these fluctuations is driven by a stimulus syllable, or a burst of intrinsic oscillations is triggered (or reset) by the sentence onset. To differentiate between these two possibilities, we compared the frequencies of the stimulus envelope and cortical responses over different compression levels; if driven by the stimulus, cortical responses to syllables should follow the syllabic frequency. But this is not the case; in this subject cortical activity exhibited two major frequency modes, one below 2 Hz and one around 5 Hz. These modes do not follow the stimulus frequency; they are stationary, albeit their relative amplitudes vary. This can also be seen in the time domain (Fig. 2A); the inter-peak intervals remained more-or-less constant (arrows) despite more than 3-fold compression of the stimulus envelope.

Frequency matching between the auditory cortex and temporal envelope

Constant-frequency behavior was observed in almost all thirteen subjects for all 3 PCs. To demonstrate this phenomenon, we present the data of all the subjects that participated in this experiment. For clarity, we divide the subjects into 3 groups according to the PC that exhibited the largest frequency mode for each subject. Data from the eight subjects in which PC1 exhibited the largest frequency mode are depicted in Figure 3. For each subject, the power spectra of the stimulus envelope and PC1 of the cortical response are plotted for each compression level (4 top rows), and her/his CI is plotted as a function of compression (lowest row). In all eight subjects, the PC1 spectra contain several modes between 0 and ~10 Hz. The main difference between responses to various speech rates was not in the location, i.e. frequency, of the modes, but rather in their relative amplitudes. In only one case (subject KR, compression 0.5), a slight shift of the ~5Hz mode towards the peak of the envelope may have occurred.

Thus, the oscillatory cortical activity does not appear to be directly driven by the syllables of the sentence. Rather, the stimulus appears to hit upon an existing set of intrinsic oscillations. In which case, what might constrain a subject's comprehension? Previously, we showed that the difference between the frequencies of the envelope and cortical mode constrains the comprehension of a subject (Ahissar et al., 2001). Careful examination of the relationships between these two variables (Fig. 3) suggests that comprehension is good when cortical frequencies are the same as, or higher than, the envelope frequencies, and becomes poorer as this relationship reverses. For example, subject RB loses comprehension already at compression of 0.5, when his cortical power drops to very low frequencies; subjects MS, KR and KB lose comprehension at compression of 0.35, when their cortical power is mainly contained in
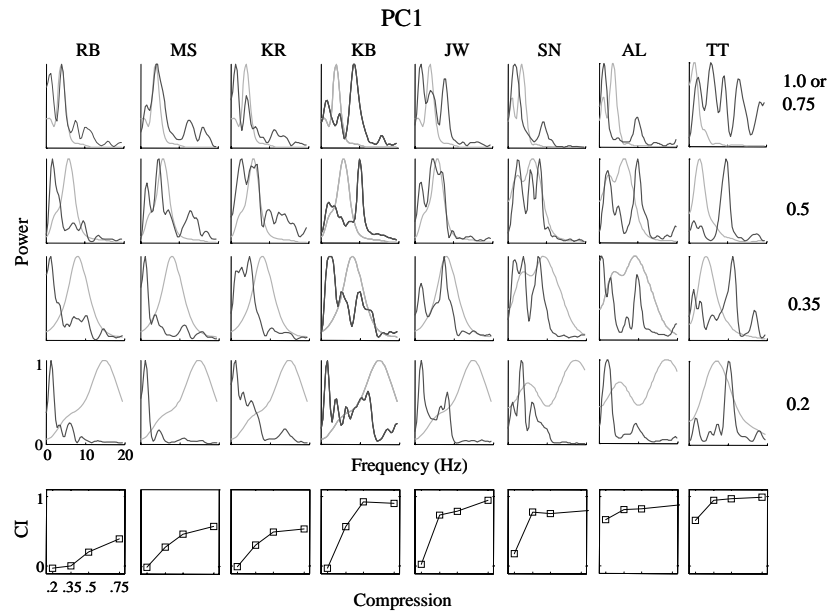
Figure 3. Power spectra and comprehension index (CI) of 8 subjects whose highest frequency mode was expressed by PC1. SN and AL were tested with compression of 1 instead of 0.75. Five subjects (left five) were tested with sentences of set 3; SN and AL with set 2, and TT with set 1. Compression levels indicated on the right.

frequencies that are lower than those of the envelope; and JW and SN lose comprehension only at compression of 0.2, when their cortical frequency becomes too low for the syllabic frequency.

These data also demonstrate that, within this range (~4 - 10 Hz), the higher the maximal intrinsic frequency generated by a cortex, the better the comprehension of the subject. This was true not only with regard to accelerated speech, but also for comprehension at normal speech rates (compressions of 0.75 or 1). In Figure 3, subjects were arranged (from left to right) according to increased comprehension (comprehension at 0.35 was arbitrarily chosen for ranking). In accordance with this gradient of comprehension, the maximal frequency generated by the left auditory cortex of these subjects, across all compression levels, also increased from left to right; from ~5 Hz with RB to ~10 Hz with the three subjects on the right. In other words, comprehension correlated with the maximal frequency that the left auditory cortex of a given subject generated when challenged with accelerated speech in our task. This arrangement revealed another signature of poor speech perceivers (in this case RB, MS and KR), they lack the ~10 Hz mode. The impairment of these three poor speech perceivers was not limited to accelerated speech, their comprehension of normal speech rates was also significantly impaired.
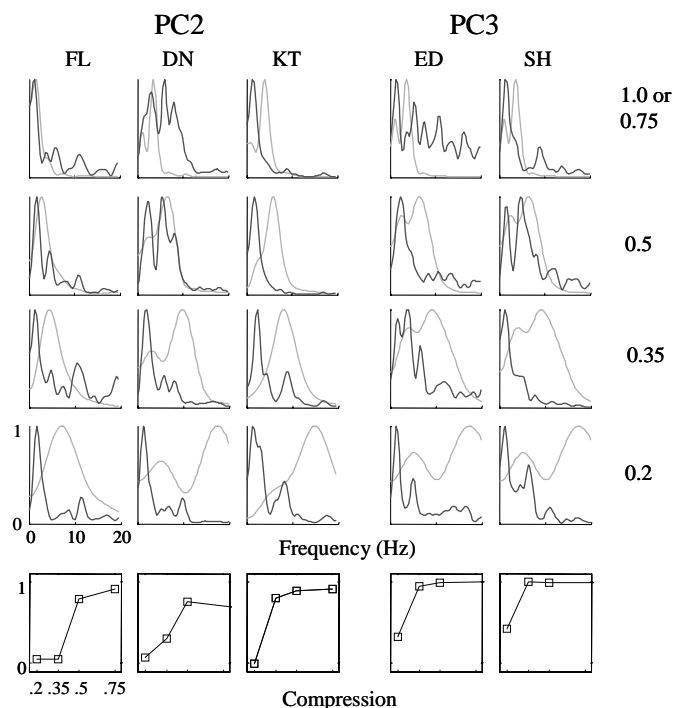
Figure 4. Power spectra and comprehension index (CI) of subjects whose highest frequency modes were expressed by PC2 (3 subjects) and PC3 (2 subjects). DN, SH and ED were tested with compression of 1 instead of 0.75. KT was tested with sentences of set 3; DN, SH and ED with set 2; and FL with set 1. Compression levels indicated on the right.

Data on the other five subjects, in which the maximal recorded frequencies were exhibited by their PC2 or PC3, are depicted in Figure 5. Although some correlation between maximal cortical frequency and comprehension was seen in these groups, it was less pronounced than that exhibited by the PC1 group (Fig. 3). For example, two subjects (KT and SH) exhibited excellent stimulus envelope and the recorded cortical responses. Subject AL also exhibited good comprehension at compression of 0.2 without such a match (Fig. 3). Perhaps these three subjects use other mechanisms or, alternatively, similar frequency-matching dependent mechanisms, but in other cortical areas, e.g., the right auditory cortex.

Interestingly, the stimulus set that had a bimodal power distribution, also often induced bimodal power distributions in cortical activity (set 2; subjects SN, AL, DN, SH, and ED). In subject ED, frequency matching with the lower mode correlated with comprehension (Fig. 4); in this subject, cortical frequency clearly tracked changes in that frequency mode as speech accelerated.

Correlation of maximal cortical frequency mode with speech comprehension and reading

Previously, we examined the correlation between comprehension of accelerated speech and stimulus-cortex frequency matching using two quantitative measures for frequency matching: (i) Fdiff (frequency difference = modal frequency of the evoked cortical signal minus the modal frequency of the stimulus envelope), and (ii) Fcc (frequency correlation coefficient = the correlation coefficient between the power spectra of the stimulus envelope and the cortical signal, in the range of 0 – 20 Hz). Both these measures of frequency matching correlate highly with comprehension: Fcc, $r = 0.87 \pm 0.12$, $p < 0.0001$; Fdiff, $r = 0.94 \pm 0.07$, $p < 0.0001$ (Ahissar et al., 2001).

We now examined the correlation between comprehension of accelerated speech and maximal frequency mode (maxFmode) in the cortical response. MaxFmode in the cortical response of a subject was defined as the highest frequency mode observed (by visual inspection of the power spectra) while a subject listened to the sentences at any of the compression levels in any of the 3 PCs. Comprehension of accelerated speech was represented as the compression "threshold" level at which comprehension was at 50% of its span (i.e., CI@50% = min(CI)+0.5[max(CI)-min(CI)]). The correlation between maxFmode and CI@50% (Fig. 5A) was apparent, but on the threshold of being statistically significant (R2=0.28, p=0.063). As the threshold level increased, correlation with maxFmode decreased (CI@25%, R2 = 0.35, p = 0.034; CI@50%, R2 = 0.28, p = 0.063; CI@75%, R2 = 0.06, p = 0.4). The correlation of maxFmode with accuracy of reading non-words ("word attack") was strong: R2 = 0.47, p = 0.027 (Fig. 5B).
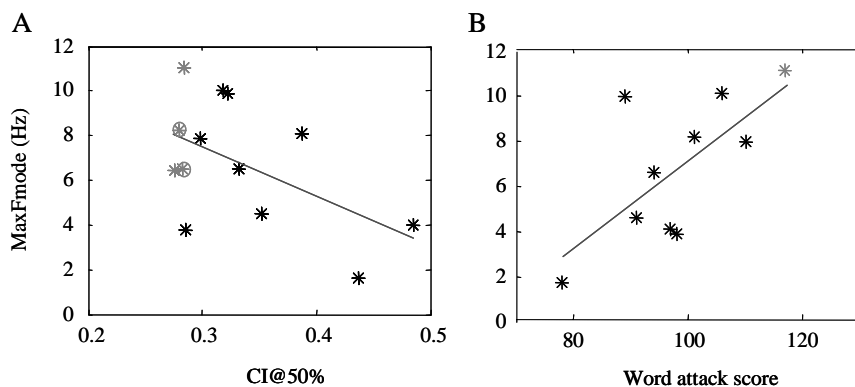

Figure 5. Correlation between the frequency of the highest mode (maxFmode) recorded from a subject and his/her comprehension threshold (CI@50%) of time-compressed speech (Panel A; $R^2 = 0.28$, p = 0.06, n=13) and accuracy of reading non-words (word attack score; Panel B; $R^2 = 0.47$, p = 0.027, n=10). RDs represented by black and good readers by gray asterisks. Data from the two non-native English speakers are circled.

The higher correlation of maxFmode with reading, than with comprehension of accelerated speech, might suggest that reading impairment and cortical rhythms do not necessarily link via mechanisms of speech perception. It is possible that both reading and speech perception depend on cognitive mechanisms related to the cortical alpha rhythm. Indeed, RDs exhibit deficiencies in perceiving successive stimuli, not only for auditory, but also for visual sequences (Hari et al., 1999a; Ben-Yehudah et al., 2001; Amitay et al., 2002b), especially in the alpha range (Ahissar M. unpublished data).

Phase locking and speech comprehension

The relevance of phase locking to speech comprehension was examined by determining the cross-correlation between the two time domain signals: the temporal envelopes of the speech input and of the recorded cortical response. The strength of phase locking was quantified as the peak-to-peak amplitude of the cross-correlation function, filtered at ±1 octave around the stimulus modal frequency at 0 to 0.5 s. This measure, which represents stimulus-response phase locking at the stimulus frequency band, correlated highly with comprehension (Ahissar et al., 2001).

The low signal-to-noise ratio of MEG signals prevented a trial-by-trial analysis. Despite this constraint, some trial specific information was obtained by comparing "correct" trials to "incorrect" and "don't know" trials. This comparison revealed that stimulus-response phase locking was significantly higher during "correct" than during "incorrect" trials or "don't know" trials, whereas frequency matching was not (Ahissar et al., 2001). This finding can be explained as follows. With short sentences, such as those presented here, presumably there is not sufficient time for the brain to change its response frequency and fully track the stimulus frequency, therefore it was crucial that the input frequency would fall within the operational range of *a priori* tuned cortical oscillating frequencies. However, the temporal patterns of the envelopes were not regular, even though they exhibited clear frequency modes. The phase-locking data indicate that comprehension was better when cortical fluctuations correlated with the pattern of the temporal envelope, not only with its average frequency. Such temporal locking could occur by chance, or due to some limited active tracking, which still occurred during these short sentences. Whatever the reason, in trials where the temporal patterns of the speech input and recorded cortical response matched, comprehension was better.

Thus, speech comprehension in these experiments appeared to depend on pre-tuned cortical oscillations whose mean frequencies were equal to or higher than the envelope frequencies, and was facilitated by accurate temporal matching. The ability of auditory cortical neurons to follow modulations at the frequency-range of speech envelopes was demonstrated in many studies (e.g., Schreiner and Urbas, 1988; Eggermont, 1998). Moreover, neurons of the primary auditory cortex of monkeys phase lock to the envelope of a portion of a species-specific vocalization, which is centered around their characteristic frequencies (Wang et al., 1995). Interestingly, the results of these studies

indicate that the strongest response locking of cortical neurons to a periodic input is usually achieved for stimulus rates (envelope frequencies) that are within the dominant range of spontaneous and evoked cortical oscillations, i.e., for frequencies below 16 Hz (Ahissar and Vaadia, 1990; Cotillon et al., 2000). This is also the frequency range in which AM modulations evoke the strongest fMRI responses in the human auditory cortex (Giraud et al., 2000). Our results suggest that cortical response locking to the temporal structure of the speech envelope is a pre-requisite for speech comprehension. This stimulus-response phase correspondence may enable internal segmentation of sentence components, mainly syllables.

Decoding of the temporal envelope: a hypothesis

Since cortical frequencies did not simply follow the input frequencies, we assume that they are generated by intrinsic mechanisms. Evidence for such intrinsic cortical mechanisms, which are capable of generating alpha range oscillations, is sound (e.g., Ahissar and Vaadia, 1990; Llinas et al., 1991; Silva et al., 1991; Gray and McCormick, 1996; Ahissar et al., 1997). The observed locking of post-stimulus cortical oscillations to speech onset (and hence not being averaged out across trials), indicates that intrinsic oscillatory sources must be triggered, or reset, by speech onset. Our results indicate that comprehension of short sentences depends on the relationship between the frequencies of these intrinsic oscillations and those of the sentence envelope: for good comprehension, cortical frequency has to be equal to or higher than the input frequency. Furthermore, our results indicate that comprehension improves when phase-locking between the sentence envelope and cortical oscillations increases.

Here, we present a hypothesis that can account for the perception of short sentences, and also continuous speech in general. Based on existing neurophysiological data and our findings, we hypothesize that arrays of intrinsic cortical oscillators, which possess frequencies in the range of natural syllabic rates (~4 to ~16 Hz), are centrally involved in speech perception. We speculate that the role of these arrays of cortical oscillators is to set the clock for spectral syllable processing (Fig. 6). Clearly, processing of the spectral content of a syllable must be done at the time in which this information is valid. Thus, the brain might extract this time from a feedforward derivative processing of syllable onset (Fishbach et al., 2001), in which case some information carried by, or close in time to, syllable onset might be lost. Alternatively, the brain could initiate syllabic processing based on an internal prediction of the time of arrival of the next syllable, in which case typically no information would be lost, and which we propose here.

We suggest that a postulated "syllabic processor" receives its timing from a cortical "temporal predictor," which is basically a bank of oscillators. These oscillators have some intrinsic frequencies that are adjustable: they can be increased by excitation and decreased by inhibition (Ahissar, 1998). Before speech is received, these oscillators can be tuned by internal processes to be in

the expected (or slightly higher) rate of speech. If tuned appropriately, the syllabic processor will be able to process the incoming syllable(s). Moreover, our data suggest that an *a priori* rough tuning, even if slightly different from the frequency of the envelopes, is sufficient for reasonable comprehension of short sentences (see Fig. 3). However, a consistent frequency difference over longer sentences should eventually impair comprehension, as cortical timing will lose input timing.

Thus, a mechanism that will force such intrinsic oscillators to lock to the actual syllabic rate is required. We propose that such a function is implemented by a mechanism similar to the phase-locked loop (PLL; Ahissar, 1998). The algorithm of the PLL is based on comparison and update (Fig. 6). Timing of the actual input envelope is compared with the predicted one, and the temporal difference is encoded by some neuronal variable, most-likely spike-count, which is used to update the period of the next oscillatory cycle. Such a
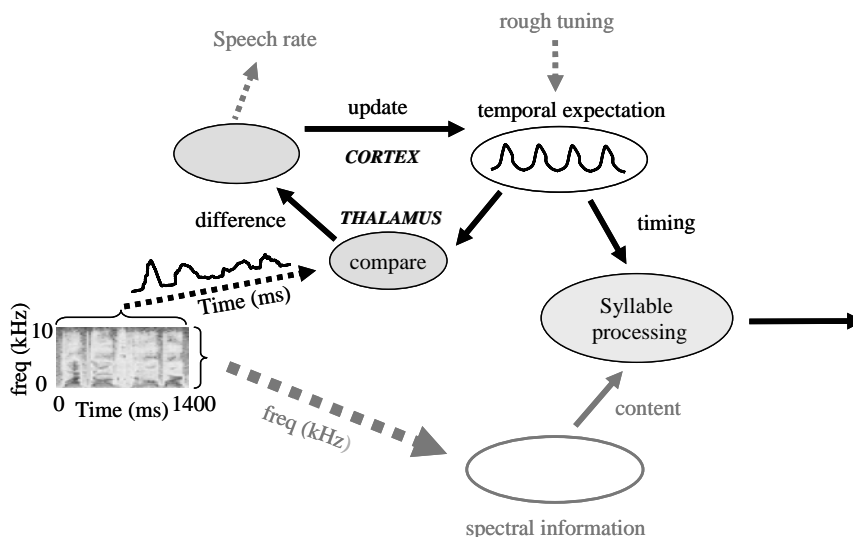

Figure 6. Hypothesized scheme for processing of the temporal envelope of speech and its function in syllable processing. We postulate that the temporal envelope of the speech signal is processed in parallel to pre-processing of its spectral content, possibly by the non-lemniscal and lemniscal systems, respectively. In this scheme, the temporal envelope is compared against a cortical reference (temporal expectation) implemented by neuronal oscillations. The comparison occurs in the (non-lemniscal) thalamus, and its output (which is proportional to the temporal difference between the two signals) is fed-back to the cortex, where it eventually updates the reference signal. The entire thalamocortical loop is a negative feedback loop that functions as a neuronal phase-locked loop (Ahissar, 1998). The signal that represents the temporal difference, which contains information about input speech rate, might be further processed (dashed arrow, top left). Cortical oscillations, which express an expectation for timing of the next syllable, send this timing information to a postulated "syllable processor," which processes dynamic spectral information contained in a syllable.

mechanism times the syllabic processor before the next syllable arrives, thus allowing efficient processing of the entire syllable, including its information-rich onset segment. Prediction of timing cannot be achieved with open-loop mechanisms, such as a bank of band-pass filters. However, a PLL, which is itself an adaptive band-pass filter, provides both efficient filtering of the actual envelope and an updated prediction about the timing of the next syllable.

The algorithm and possible neuronal implementations of such a loop were previously discussed in detail (Ahissar and Vaadia, 1990; Ahissar et al., 1997; Ahissar, 1998; Kleinfeld et al., 1999; Ahissar and Arieli, 2001; Ahissar and Zacksenhouse, 2001; Kleinfeld et al., 2002), and thus, are not described here. Since auditory implementations of the PLL algorithm have not been proposed before, we will suggest auditory sites that appear plausible for such implementation. Our suggestions are based on known auditory anatomy and physiology, and on an analogy with the tactile system of the rat, in which some predictions of the PLL algorithm were recently tested and confirmed (Ahissar et al., 1997; Kleinfeld et al., 1999; Ahissar et al., 2000a; Ahissar and Arieli, 2001; Sosnik et al., 2001).

In principle, PLLs might be implemented by many neuronal circuits, including sub-cortical and cortico-cortical loops. We hypothesize that envelope-related PLLs, like those in the tactile system, are implemented by thalamocortical loops, and that a temporal comparison takes place in the thalamus, most likely the non-lemniscal nuclei. As in the vibrissal tactile system, the auditory non-lemniscal thalamus exhibits larger temporal dispersion (He and Hu, 2002; Hu, 2003) and spatial (spectral) integration (Steriade et al., 1997; Malmierca et al., 2002; Hu, 2003) than the lemniscal thalamus. Both these features, together with typical thalamic gating mechanisms (McCormick and von Krosigk, 1992; Sherman and Guillery, 1996) and strong corticothalamic feedback (Ojima, 1994), make the non-lemniscal thalamus optimal for temporal comparison functions (Ahissar, 1998; Ahissar and Arieli, 2001). According to this hypothesis, the signal indicating timing differences, produced by the non-lemniscal thalamus, is fed back to the cortex where it is used to update the frequency of the intrinsic oscillators. The entire loop should be connected as a negative feedback loop, via thalamic (Yu et al., 2004) or cortical inhibition. Such a negative feedback loop would force the cortical oscillations to phase-lock to the envelope of the speech signal. Thus, with longer spoken sentences we predict that cortical oscillations will not stick to their original frequencies, but will adapt to the actual envelope frequencies.

Spectral analysis of syllables must be coordinated with syllable timing. In fact, Giraud et al. (2000) suggested that "Slow oscillations (2 – 10 Hz) could signal the temporal limits of syllables." Our hypothesis, which is fully consistent with this, goes further by suggesting that these slow oscillations are intrinsic, and actually serve as temporal expectations. It is the intrinsic "clock" that sets the pace for segmentation, and not the incoming signal. The rate of the incoming signal only updates the intrinsic clock, and makes it a better predictor of the following input rate. The existence of such an intrinsic clock is consistent with another phenomenon often related to speech perception, that of perceptual center (P-center; Morton et al., 1976; Goswami et al., 2002). The P-center of a

signal, such as a syllable, "corresponds to its psychological moment of occurrence" (Morton et al., 1976). The experimental definition of a P-center of a speech signal depends on a comparison, by a listener, between such psychological moments of occurrence and an internal "temporal ruler," or pacemaker (Morton et al., 1976; Scott, 1993). Our model describes such a process: the rate of an input stream of syllables is compared against an internal temporal ruler, in the form of intrinsic oscillations.

Summary and conclusions

We have extended our previous report on processing of the temporal envelope of speech by demonstrating the difference between poor and good readers in comprehending accelerated speech, and demonstrating the correlation of the maxFmode recorded from the auditory cortex with reading and with comprehension thresholds for accelerated speech. Our hypothesis about the processing of speech rate, and its possible role in speech comprehension, is described in detail. Based on our hypothesis, we suggest that the ability of listeners to adapt to varying speech rates depends on the dynamic range of their cortical oscillations. Further experimental testing will determine whether these dynamic ranges can be increased by training, and whether such increases will facilitate comprehension of varying-rate speech.

Acknowledgments

References

Ahissar, E. (1998). Temporal-code to rate-code conversion by neuronal phase-locked loops. Neural Computation, 10, 597-650.

Ahissar, E. & Vaadia, E. (1990). Oscillatory activity of single units in a somatosensory cortex of an awake monkey and their possible role in texture analysis. Proceedings of the National Academy of Science USA, 87, 8935-8939.

Ahissar, E. & Zacksenhouse, M. (2001). Temporal and spatial coding in the rat vibrissal system. Progress in Brain Research, 130, 75-88.

Ahissar, E. & Arieli, A. (2001). Figuring space by time. Neuron, 32, 185-201.

Ahissar, E., Haidarliu, S. & Zacksenhouse, M. (1997). Decoding temporally encoded sensory input by cortical oscillations and thalamic phase comparators. Proceedings of the National Academy of Science USA, 94, 11633-11638.

Ahissar, E., Sosnik, R. & Haidarliu, S. (2000a). Transformation from temporal to rate coding in a somatosensory thalamocortical pathway. Nature, 406, 302-306.

Ahissar, E., Nagarajan, S., Ahissar, M., Protopapas, A., Mahncke, H. & Merzenich, M. (2001). Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. Proceedings of the National Academy of Science USA, 98, 13367-13372.

Ahissar, M., Protopapas, A., Reid, M. & Merzenich, M.M. (2000b). Auditory processing parallels reading abilities in adults. Proceedings of the National Academy of Science USA, 97, 6832-6837.

Amitay, S., Ahissar, M. & Nelken, I. (2002a). Auditory processing deficits in reading disabled adults. Journal of the Association for Research in Otolaryngology, 3, 302-320.

Amitay, S., Ben-Yehudah, G., Banai, K. & Ahissar, M. (2002b). Disabled readers suffer from visual and auditory impairments but not from a specific magnocellular deficit. Brain, 125, 2272-2285.

Aram, D.M., Ekelman, B.L. & Nation, J.E. (1984). Preschoolers with language disorders: 10 years later. Journal of Speech & Hearing Research, 27, 232-244.

Banai, K. & Ahissar, M. (in press). Psychoacoustics and working memory in dyslexia. In J. Syka & M.M. Merzenich, (Eds.), Proceedings of the symposium on plasticity of the central auditory system and processing of complex acoustic signals. New York: Kluwer Academic/Plenum Publishers.

Beasley, D.S., Bratt, G.W. & Rintelmann, W.F. (1980). Intelligibility of time-compressed sentential stimuli. Journal of Speech & Hearing Research, 23, 722-731.

Ben-Yehudah, G., Sackett, E., Malchi-Ginzberg, L. & Ahissar, M. (2001). Impaired temporal contrast sensitivity in dyslexics is specific to retain-and-compare paradigms. Brain, 124, 1381-1395.

Bishop, D.V.M. (1992). The underlying nature of specific language impairment. Journal of Child Psychology & Psychiatry & Allied Disciplines, 33, 2-66.

Cotillon, N., Nafati, M. & Edeline, J.M. (2000). Characteristics of reliable tone-evoked oscillations in the rat thalamo- cortical auditory system. Hearing Research, 142, 113-130.

Drullman, R., Festen, J.M. & Plomp, R. (1994). Effect of temporal envelope smearing on speech reception. Journal of the Acoustical Society of America, 95, 1053-1064.

Dupoux, E. & Green, K. (1997). Perceptual adjustment to highly compressed speech: effects of talker and rate changes. Journal of Experimental Psychology - Human Perception and Performance, 23, 914-927.

Eggermont, J.J. (1998). Representation of spectral and temporal sound features in three cortical fields of the cat. Similarities outweigh differences. Journal of Neurophysiology, 80, 2743-2764.

Farmer, M.E. & Klein, R.M. (1995). The evidence for a temporal processing deficit linked to dyslexia: A review. Psychonomics Bulletin & Review, 2, 460-493.

Fishbach, A., Nelken, I. & Yeshurun, Y. (2001). Auditory edge detection: a neural model for physiological and psychoacoustical responses to amplitude transients. Journal of Neurophysiology, 85, 2303-2323.

Foulke, E. & Sticht, T.G. (1969). Review of research on the intelligibility and comprehension of accelerated speech. Psychological Bulletin, 72, 50-62.

Freeman, B.A. & Beasley, D.S. (1978). Discrimination of time-altered sentential approximations and monosyllables by children with reading problems. Journal of Speech & Hearing Research, 21, 497-506.

Giraud, A.L., Lorenzi, C., Ashburner, J., Wable, J., Johnsrude, I., Frackowiak, R. & Kleinschmidt, A. (2000). Representation of the temporal envelope of sounds in the human brain. Journal of Neurophysiology, 84, 1588-1598.

Goswami, U., Thomson, J., Richardson, U., Stainthorp, R., Hughes, D., Rosen, S. & Scott, S.K. (2002). Amplitude envelope onsets and developmental dyslexia: A new hypothesis. Proceedings of the National Academy of Science USA, 99, 10911-10916.

Gray, C.M. & McCormick, D.A. (1996). Chattering cells: superficial pyramidal neurons contributing to the generation of synchronous oscillations in the visual cortex. Science, 274, 109-113.

Hari, R., Valta, M. & Uutela, K. (1999a). Prolonged attentional dwell time in dyslexic adults. Neuroscience Letters, 271, 202-204.

Hari, R., Saaskilahti, A., Helenius, P. & Uutela, K. (1999b). Non-impaired auditory phase locking in dyslexic adults. Neuroreport, 10, 2347-2348.

He, J. & Hu, B. (2002). Differential distribution of burst and single-spike responses in auditory thalamus. Journal of Neurophysiology, 88, 2152-2156.

Helenius, P., Uutela, K. & Hari, R. (1999). Auditory stream segregation in dyslexic adults. Brain, 122 (Pt 5), 907-913.

Houtgast, T. & Steeneken, H.J.M. (1985). A review of the MFT concept in room acoustics and its use for estimating speech intelligibility in auditoria. Journal of the Acoustical Society of America, 77, 1069-1077.

Hu, B. (2003). Functional organization of lemniscal and nonlemniscal auditory thalamus. Experimental Brain Research, 153, 543-549.

Kleinfeld, D., Berg, R.W. & O'Connor, S.M. (1999). Anatomical loops and their electrical dynamics in relation to whisking by rat. Somatosensory & Motor Research, 16, 69-88.

Kleinfeld, D., Sachdev, R.N., Merchant, L.M., Jarvis, M.R. & Ebner, F F. (2002). Adaptive filtering of vibrissa input in motor cortex of rat. Neuron, 34, 1021-1034.

Llinas, R.R., Grace, A.A. & Yarom, Y. (1991). In vitro neurons in mammalian cortical layer 4 exhibit intrinsic oscillatory activity in the 10- to 50-Hz frequency range. Proceedings of the National Academy of Science USA, 88, 897-901.

Malmierca, M.S., Merchan, M.A., Henkel, C.K. & Oliver, D.L. (2002). Direct projections from cochlear nuclear complex to auditory thalamus in the rat. Journal of Neuroscience, 22, 10891-10897.

McAnally, K.I., Hansen, P.C., Cornelissen, P.L. & Stein, J.F. (1997). Effect of time and frequency manipulation on syllable perception in developmental dyslexics. Journal of Speech Language & Hearing Research, 40, 912-924.

McCormick, D.A. & von Krosigk, M. (1992). Corticothalamic activation modulates thalamic firing through glutamate "metabotropic" receptors. Proceedings of the National Academy of Science USA, 89, 2774-2778.

Miller, J.L., Grosjean, F. & Lomanto, C. (1984). Articulation rate and its variability in spontaneous speech: a reanalysis and some implications. Phonetica, 41, 215-225.

Morton, J., Marcus, S. & Framkish, C. (1976). Perceptual centers (P-centers). Psychological Review, 83, 405-408.

Newman, R.S. & Sawusch, J.R. (1996). Perceptual normalization for speaking rate: effects of temporal distance. Perception & Psychophysics, 58, 540-560.

Ojima, H. (1994). Terminal morphology and distribution of corticothalamic fibers originating from layers 5 and 6 of cat primary auditory cortex. Cerebral Cortex, 4, 646-663.

Renvall, H. & Hari, R. (2002). Auditory cortical responses to speech-like stimuli in dyslexic adults. Journal of Cognitive Neuroscience, 14, 757-768.

Riensche, L.L. & Clauser, P.S. (1982). Auditory perceptual abilities of formerly misarticulating children. Journal of Auditory Research, 22, 240-248.

Rosen, S. (1992). Temporal information in speech: acoustic, auditory and linguistic aspects. Philosophical Transactions of the Royal Society of London Series B - Biological Science, 336, 367-373.

Schreiner, C.E. & Urbas, J.V. (1988). Representation of amplitude modulation in the auditory cortex of the cat. II. Comparison between cortical fields. Hearing Research, 32, 49-63.

Scott, S.K. (1993). P-centers - an acoustic analysis. PhD thesis, University College London.

Shapiro, K.L., Ogden, N. & Lind-Blad, F. (1990). Temporal processing in dyslexia. Journal of Learning Disabilities, 23, 99-107.

Sherman, S.M. & Guillery, R.W. (1996). Functional organization of thalamocortical relays. Journal of Neurophysiology, 76, 1367-1395.

Silva, L.R., Amitai, Y. & Connors, B.W. (1991). Intrinsic oscillations of neocortex generated by layer 5 pyramidal neurons. Science, 251, 432-435.

Sosnik, R., Haidarliu, S. & Ahissar, E. (2001). Temporal frequency of whisker movement. I. Representations in brain stem and thalamus. Journal of Neurophysiology, 86, 339-353.

Steriade, M., Jones, E.G. & McCormick, D.A. (1997). Thalamus. Vol. I: Organisation and function. Amsterdam: Elsevier.

Tallal, P. & Piercy, M. (1973). Defects of non-verbal auditory perception in children with developmental aphasia. Nature, 241, 468-469.

Tallal, P., Miller, S. & Fitch, R.H. (1993). Neurobiological basis of speech: a case for the preeminence of temporal processing. Annals of the NY Academy of Science, 682, 27-47.

van der Horst, R., Leeuw, A.R. & Dreschler, W.A. (1999). Importance of temporal-envelope cues in consonant recognition. Journal of the Acoustical Society of America, 105, 1801-1809.

Wang, X., Merzenich, M.M., Beitel, R. & Schreiner, C.E. (1995) Representation of a species-specific vocalization in the primary auditory cortex of the common marmoset: Temporal and spatial characteristics. Journal of Neurophysiology, 74, 2685-2706.

Watson, M., Stewart, M., Krause, K. & Rastatter, M. (1990). Identification of time-compressed sentential stimuli by good vs poor readers. Perceptual Motor Skills, 71, 107-114.

Welsh, L.W., Welsh, J.J., Healy, M. & Cooper, B. (1982). Cortical, subcortical, and brainstem dysfunction: a correlation in dyslexic children. Annals of Otology Rhinology & Laryngology, 91, 310-315.

Woodcock, R. (1987). Woodcock Reading Mastery Tests - Revised. Circle Pines, MN: American Guidance Service.

Yu, Y., Xiong, Y., Chan, Y. & He, J. (2004). Corticofugal gating of auditory information in the thalamus: An in vivo intracellular recording study. Journal of Neuroscience 24:3060-3069.