



# Speech comprehension is correlated with temporal response patterns recorded from auditory cortex

Ehud Ahissar<sup>\*†</sup>, Srikantan Nagarajan<sup>‡</sup>, Merav Ahissar<sup>§</sup>, Athanassios Protopapas<sup>¶||</sup>, Henry Mahncke<sup>\*\*</sup>, and Michael M. Merzenich<sup>||\*\*</sup>

<sup>\*</sup>Department of Neurobiology, Weizmann Institute of Science, Rehovot 76100, Israel; <sup>†</sup>Department of Bioengineering, University of Utah, Salt Lake City, UT, 84112; <sup>‡</sup>Department of Psychology, Hebrew University, Jerusalem 91905, Israel; <sup>§</sup>Department of Speech Technology, Institute for Language and Speech Processing, 15125 Maroussi, Greece; <sup>\*\*</sup>The Keck Center for Integrative Neurosciences, University of California, San Francisco, CA 94143-0732; and <sup>||</sup>Scientific Learning Corporation, 300 Frank H. Ogawa Plaza, Suite 500, Oakland, CA 94612-2040

Contributed by Michael M. Merzenich, July 31, 2001

**Speech comprehension depends on the integrity of both the spectral content and temporal envelope of the speech signal. Although neural processing underlying spectral analysis has been intensively studied, less is known about the processing of temporal information. Most of speech information conveyed by the temporal envelope is confined to frequencies below 16 Hz, frequencies that roughly match spontaneous and evoked modulation rates of primary auditory cortex neurons. To test the importance of cortical modulation rates for speech processing, we manipulated the frequency of the temporal envelope of speech sentences and tested the effect on both speech comprehension and cortical activity. Magnetoencephalographic signals from the auditory cortices of human subjects were recorded while they were performing a speech comprehension task. The test sentences used in this task were compressed in time. Speech comprehension was degraded when sentence stimuli were presented in more rapid (more compressed) forms. We found that the average comprehension level, at each compression, correlated with (i) the similarity between the frequencies of the temporal envelopes of the stimulus and the subject's cortical activity ("stimulus-cortex frequency-matching") and (ii) the phase-locking (PL) between the two temporal envelopes ("stimulus-cortex PL"). Of these two correlates, PL was significantly more indicative for single-trial success. Our results suggest that the match between the speech rate and the *a priori* modulation capacities of the auditory cortex is a prerequisite for comprehension. However, this is not sufficient: stimulus-cortex PL should be achieved during actual sentence presentation.**

human | MEG | time compression | accelerated speech | phase-locking

Comprehension of speech depends on the integrity of its temporal envelope, that is, on the temporal variations of spectral energy. The temporal envelope contains information that is essential for the identification of phonemes, syllables, words, and sentences (1). Envelope frequencies of normal speech are usually below 8 Hz (ref. 2; see Figs. 1 and 2). The critical frequency band of the temporal envelope for normal speech comprehension is between 4 and 16 Hz (3, 4); envelope details above 16 Hz have only a small [although significant (5)] effect on comprehension. Across this low-frequency modulation range, comprehension does not usually depend on the exact frequencies of the temporal envelopes of incoming speech, because the temporal envelope of normal speech can be compressed in time down to 0.5 of its original duration before comprehension is significantly affected (6, 7). Thus, normal brain mechanisms responsible for speech perception can adapt to different input rates within this range (see refs. 8–10). This online adaptation is crucial for speech perception, because speech rates vary between different speakers and change according to the emotional state of the speaker.

Poor readers, many of whom have poor successive-signal auditory (11–17) and visual (18) processing, are more vulnerable than good readers to the time compression of sentences (19–21),

although not to speech compression of syllables (22). Comparison of evoked responses suggests that the deficiencies of poor readers at tasks requiring the recognition of time-compressed speech emerge at the cortical level (23). Taken together, these findings suggest that the auditory cortex can process speech sentences at various rates, but that the extent of the "decodable ranges" of speech modulation rates can substantially vary from one listener to another. More specifically, the ranges of poor readers seem to be narrower, and shifted downward, than those of good readers.

Over the past decade, several magnetoencephalographic (MEG) studies have shown that magnetic field signals arising from the primary auditory cortex and surrounding cortical areas on the superior temporal plane can provide valuable information about the spectral and temporal processing of speech stimuli (24–27). The magnetoencephalogram (MEG) is currently the most suitable noninvasive technology for accurately measuring the dynamics of neural activity within specific cortical areas, especially on the millisecond time scale. It has been shown previously that the perceptual identification of ordered non-speech acoustic stimuli is correlated with aspects of auditory MEG signals (28–30). Here, we were interested in examining possible neuronal correlates for speech perception. More specifically, we asked whether the behavioral dependence of speech comprehension on the speech rate is paralleled by a similar behavior of appropriate aspects of neuronal activity located to the general area of the primary auditory cortical field. Toward that end, MEG signals arising from the auditory cortices were recorded in human subjects (Ss) while they were processing speech sentences at four different time compressions. Ss for this study were selected from a population with a wide spectrum of reading abilities, to cover a large range of competencies in their effective processing of accelerated speech.

## Methods

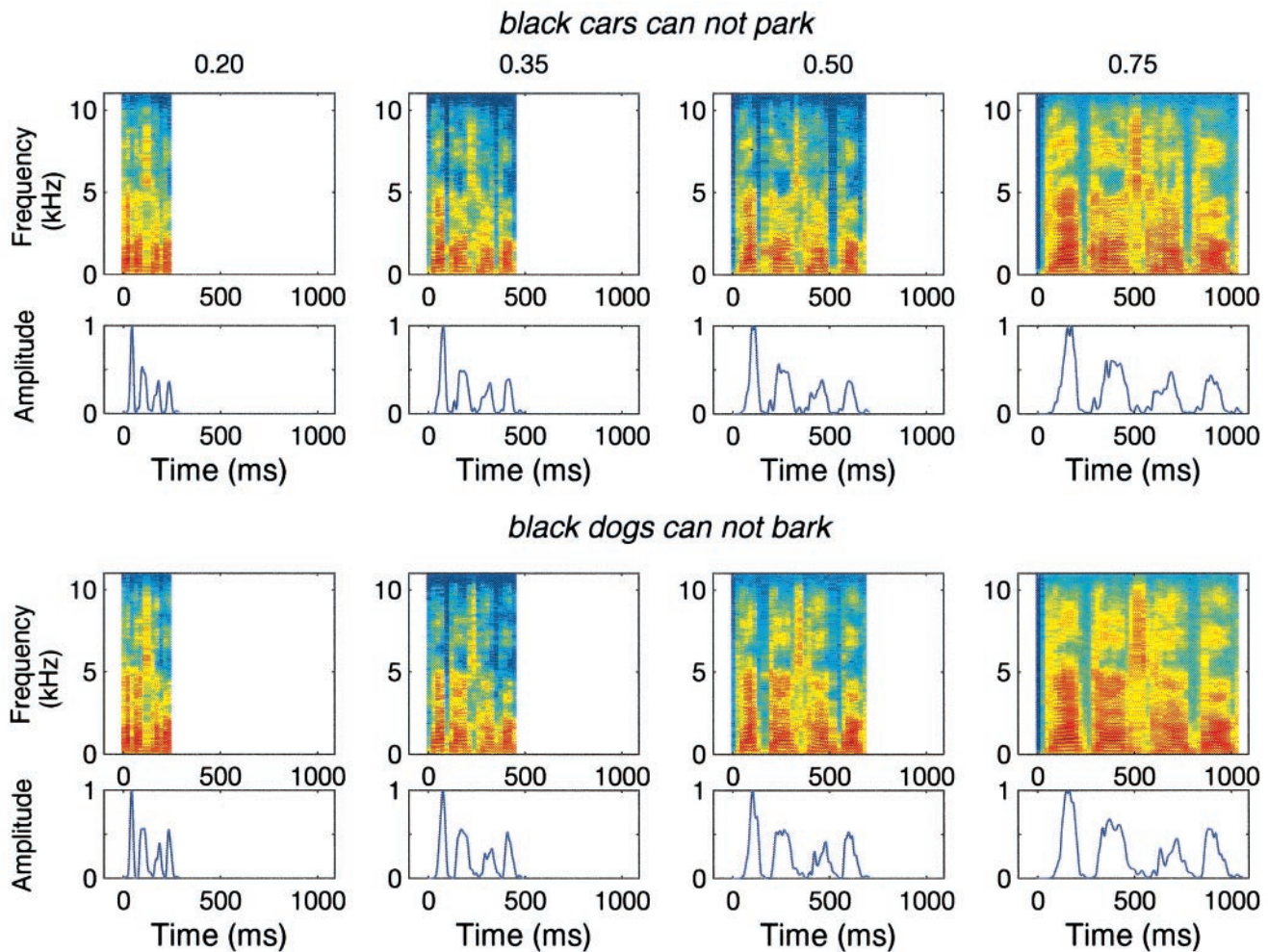
**Subjects.** Thirteen Ss (7 males and 6 females, aged 25–45) volunteered to participate in the experiment. Reading abilities spanned the ranges of 81–122 in a word-reading test, and 78–117 in a nonword reading test (31). Eleven Ss were native English speakers; two used English as their second language. All participants gave their written informed consent for the behavioral and MEG parts of the study. Studies were performed with the approval of an institutional committee for human research.

**Acoustic Stimuli.** Before the speech comprehension experiment, 1-kHz tone pips, which were 400 ms in total duration with 5-ms

Abbreviations: MEG, magnetoencephalogram/magnetoencephalographic; PC, principal component; Fdiff, frequency difference; PL, phase-locking; S, subject; MUSIC, multiple signal classification; Fcc, frequency correlation coefficient.

<sup>†</sup>To whom reprint requests should be addressed. E-mail: Ehud.Ahissar@weizmann.ac.il.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.



**Fig. 1.** Compressed speech stimuli. Shown here are two sample sentences used in the experiment. Rows 1 and 3 show the spectrogram of the sentences “black cars cannot park” and “black dogs cannot bark,” respectively. Rows 2 and 4 show the corresponding temporal envelopes of these sentences. Columns correspond to compression ratios of (left to right) 0.2, 0.35, 0.5, and 0.75.

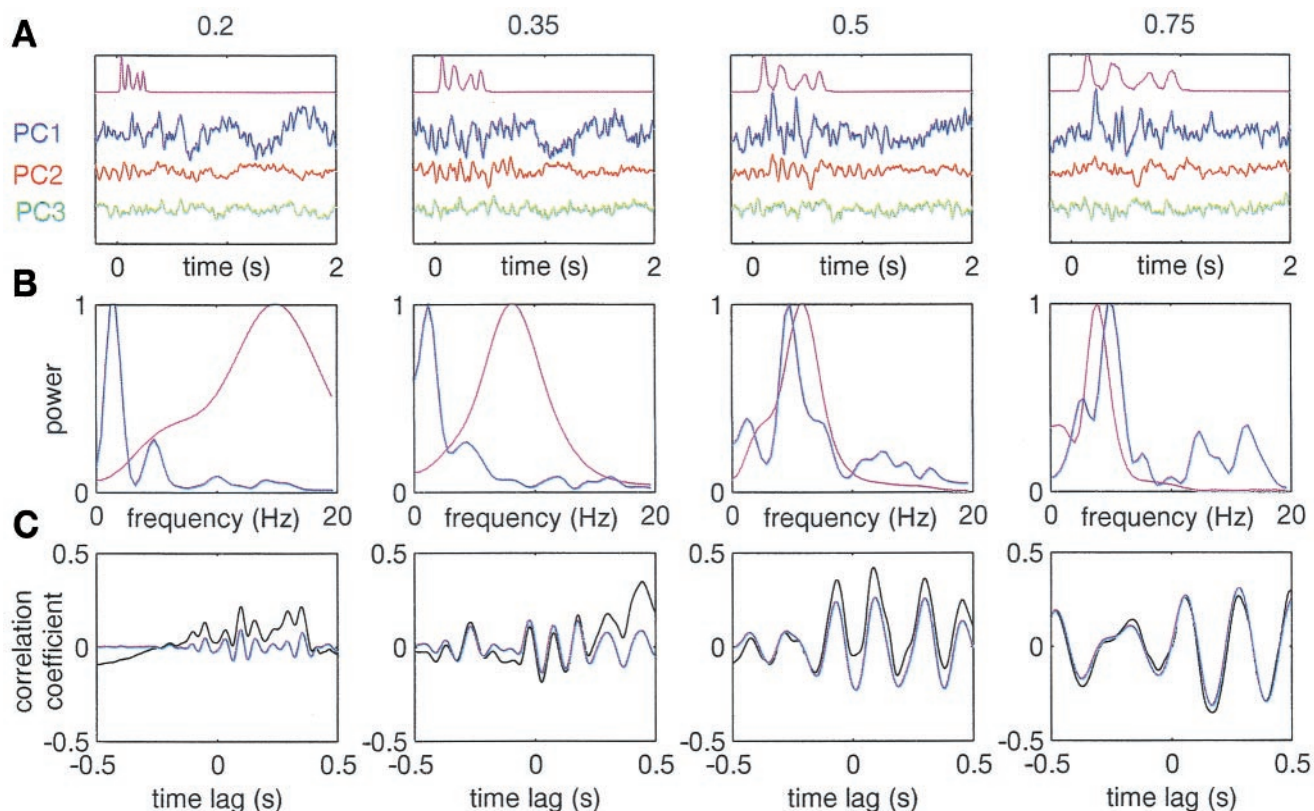
rise and fall ramps and presented at 90 dB sound pressure level (SPL) in amplitude, were used to optimize the position of the MEG magnetic signal recording array over the auditory cortex. For the compressed speech comprehension experiment, a list of several sentences uttered at a natural speaking rate was first recorded digitally from a single female speaker. Then sentences were compressed to different rates by applying a time-scale compression algorithm that kept the spectral and pitch content intact across different compression ratios. The time-scale algorithm used was based on a modified form of a phase-vocoder algorithm (32) and produced artifact-free compression of the speech sentences (Fig. 1). Onsets were aligned for different sentences and compressions, with data acquisition triggered on a pulse-marking sentence onset. Stimulus delivery was controlled by a program written in LABVIEW (National Instruments, Austin, TX). Sentence stimuli were delivered through an Audiomeia card (Digidesign, Palo Alto, CA) at conversation levels of  $\approx 70$  dB SPL.

**Sentences.** Three balanced sets of sentences were used. Set 1 included four different sentences. “Two plus six equals nine.” “Two plus three equals five.” “Three plus six equals nine.” “Three plus three equals five.” Set 2 also included four different sentences. “Two minus two equals none.” “Two minus one equals one.” “Two minus two equals one.” “Two minus one

equals none.” Set 3 included 10 sentences. “Black cars can all park.” “Black cars cannot park.” “Black dogs can all bark.” “Black dogs cannot bark.” “Black cars can all bark.” “Black cars cannot bark.” “Black dogs can all park.” “Black dogs cannot bark.” “Playing cards can all park.” “Playing cards cannot park.” Each S was tested with sentences from one set. The sentences in each set were selected so that (i) there were an equal number of true and false sentences, (ii) there were no single words on which Ss could base their answers, and (iii) the temporal envelopes for different sentences were similar. Correlation coefficients between single envelopes and the average envelope were (mean  $\pm$  SD)  $0.71 \pm 0.14$  for set 1,  $0.82 \pm 0.04$  for set 2, and  $0.91 \pm 0.07$  for set 3.

**Experiment.** Ss were presented with sentences at compression ratios (compressed sentence duration/original sentence duration) of 0.2, 0.35, 0.5, and 0.75. For each sentence, Ss responded by pressing one of three buttons corresponding to “true,” “false,” or “don’t know,” signaling answers by using their left hands. Compression ratios and sentences were balanced and randomized. A single psychophysical/imaging experiment typically lasted for about 2 h.

**Recordings.** Magnetic fields were recorded from the left hemisphere in a magnetically shielded room by using a 37-channel



**Fig. 2.** An example of MEG signals recorded during the task, and the measures derived from them (5 MS). (A) Averaged temporal envelopes (magenta) and the first three PCs (PC1–3, blue, red, and green, respectively, scaled in proportion to their eigen values) of the averaged responses. (B) Power spectra of the stimulus envelope (magenta) and PC1 (blue). (C) Time domain cross correlation between the envelope and PC1; black, raw correlation; blue, after band-pass filtering at  $\pm 1$  octave around the stimulus modal frequency.

biomagnetometer array with superconducting *quantum interference device* (SQUID)-based first-order gradiometer sensors (Magnes II; Biomagnetic Technology, San Diego). Fiducial points were marked on the skin for later coregistration with structural MRIs, and the head shape was digitized to constrain subsequent source modeling. The sensor array was initially positioned over an estimated location of auditory cortex in the left hemisphere so that a dipolar response was evoked by single 400-ms tone pips. Data acquisition epochs were 600 ms in total duration with a 100-ms prestimulus period referenced to the onset of the tone sequence. Data were acquired at a sampling rate of 1,041 Hz. Then the position of the sensor was refined so that a single dipole localization model resulted in a correlation and goodness of fit greater than 95% for an averaged evoked magnetic field response to 100 tones. After satisfactory sensor positioning over the auditory cortex, Ss were presented with sentences at different compression ratios. Data acquisition epochs were 3,000 ms in total duration with a 1,000-ms prestimulus period. Data were acquired at a sampling rate of 297.6 Hz.

**Data Analysis.** For each S, data were first averaged across all artifact-free trials. Then 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) were calculated. The PCs typically accounted for more than 90% of the variance within the sensor array. The eigen vectors and values obtained from the averaged data were used for all computations related to that S. Then data were divided into categories according to compression ratio and response class (“correct,” “incorrect,” and don’t know). Trials were

averaged, and the first three PCs were computed for each class (Fig. 2A). The following measures were derived from the 2-s poststimulus period by computing and averaging measures for each PC weighted by its eigen value. (i) rms of the cortical signal. (ii) frequency difference (Fdiff) = modal frequency of the evoked cortical signal minus the modal frequency of the stimulus envelope. Modal frequencies (i.e., frequencies of maximal power) were computed from the fast Fourier transforms (FFTs) of the envelope and signals (see Fig. 2B). FFTs were computed by using windows of 1 s and overlaps of 0.5 s. (iii) Frequency correlation coefficient (Fcc) = the correlation coefficient between the FFT of the stimulus envelope and the FFT of the cortical signal, in the range of 0–20 Hz. (iv) Phase-locking (PL) = peak to peak amplitude of the temporal cross correlation between the stimulus envelope and the cortical signal within the range of time lags 0–0.5 s. The cross correlation was first filtered by a band-pass filter at  $\pm 1$  octave around the modal frequency of the stimulus envelope (see Fig. 2C). Dependencies of these average measures on the compression ratio and response type were correlated with speech comprehension. Comprehension was quantified as  $C = (n \text{ correct} - n \text{ incorrect})/n \text{ trials}$ .  $C$  could have values between  $-1$  (all incorrect) and  $1$  (all correct), where  $0$  was the chance level.

**Multiple Dipole Localization.** Multiple dipole localization analyses of spatiotemporal evoked magnetic fields were performed by using an algorithm called *multiple signal classification* (MUSIC; ref. 33). MUSIC methods are based on estimation of a signal “subspace” from entire spatiotemporal MEG data by using singular-value decomposition (SVD). A version of the

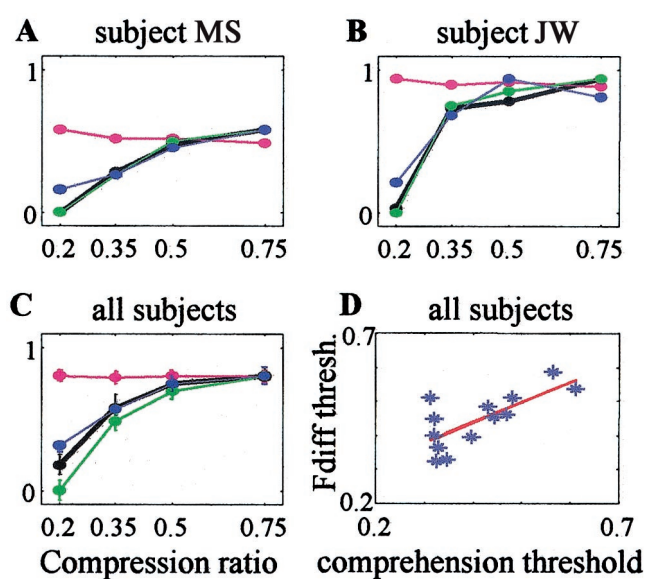
MUSIC algorithm, referred to as “the conventional MUSIC algorithm,” was implemented in MATLAB (Math Works, Natick, MA) under the assumption that the sources contributing to the MEG data arose from multiple stationary dipoles (<37 in number) located within a spherical volume of uniform conductivity (34). The locations of dipoles are typically determined by conducting a search over a three-dimensional (3D) grid of interest within the head. Given the sensor positions and the coordinates of the origin of a “local sphere” approximation of the head shape for each S, a Lead-field matrix was computed for each point in this 3D grid. From these Lead-field matrices and the covariance matrices of spatiotemporal MEG data, the value of a MUSIC localizer function could be computed (equation 4 in ref. 34). Minima of this localizer function correspond to the location of dipolar sources. For each S, at each point in a 3D grid ( $-4 < x < 6$ ,  $0 < y < 8$ ,  $3 < z < 11$ ) in the left hemisphere, the localizer function was computed over a period after sentence onset by using the averaged evoked auditory magnetic field responses.

## Results

At the beginning of each recording session, sensor array location was adjusted to yield an optimal MEG signal across the 37 channels (see *Methods*). To confirm that the location of the source dipole(s) was within the auditory cortex, the MUSIC algorithm was run on recorded responses to test sentences. For all Ss, it yielded a single dipole source. The exact location of the peaks of these localizer functions varied across Ss according to their head geometries and the locations of their lateral fissure and superior temporal sulci. However, for all Ss, the locations of minima were within 2–3 mm of the average coordinates of the primary auditory cortical field on Heschl’s gyrus (0.5, 5.1, 5.0 cm; refs. 35 and 36). When these single dipoles were superimposed on three-dimensional structural MRIs, they were invariably found to be located on the supratemporal plane, approximately on Heschl’s gyrus.

The low signal to noise ratio of MEG recordings requires data averaged across multiple repetitions of the same stimuli. This requirement imposed a practical limit on the number of sentences that could be used. To reduce a possible dependency of results on a specific stimulus set, we used three contextually different sets of sentences. Sentences in each set were designed to yield similar temporal envelopes so that trials of different sentences with the same compression ratios could be averaged to improve signal to noise ratio. PC analyses conducted on such averaged data were used to reduce the dimensionality of the MEG sensor array from 37 to 3 while preserving most of the information contained in the MEG signals (Fig. 2A; see *Methods*).

To examine the extent of frequency-matching between the temporal envelope of the stimulus and that of recorded MEG signals, power spectra of the stimulus envelope and the three PCs were computed (Fig. 2B; only PC1 is shown). The modal frequency of evoked cortical signals was fairly close to that of the stimulus for compression ratios of 0.75 and 0.5 (also compare the time-domain signals, Fig. 2A). However, for stronger compressions, the frequency of the cortical signals could not follow the speech signal modulation, and the difference between the modal frequencies of the stimulus and the cortical signals progressively increased. The difference between modal frequencies of the stimulus vs. auditory cortex responses (“Fdiff”; see *Methods*) was correlated with sentence comprehension (“C”; see *Methods*). For S MS shown in Figs. 2 and 3A for example, Fdiff (green curve) and comprehension (black curve) were strongly correlated ( $P = 0.002$ , linear regression analysis), as is demonstrated by the overlap of the normalized curves. In fact, Fdiff and C were significantly correlated ( $P < 0.05$ ) in 10 of 13 Ss (see another example in Fig. 3B). On the average, Fdiff could predict



**Fig. 3.** Neuronal correlates for speech comprehension. A–C measures were averaged across PC1–3 (see *Methods*) and normalized to the maximal value of the comprehension curve; neuronal correlates with negative values were first “shifted” up by adding a constant so that their minimal value became 0. (A and B) Comprehension (black thick curve) and neuronal correlates (magenta, rms; green, Fdiff; blue, PL) for the S depicted in Fig. 2 (MS) and for another S (JW). (C) Average comprehension and neuronal correlates across all Ss (mean  $\pm$  SEM,  $n = 13$ ). (D) Scatter plot of thresholds for comprehension and Fdiff for all Ss. For each variable and each S, threshold was the (interpolated) compression ratio corresponding to 0.75 of the range spanned by that variable. The red line is a linear regression.

88% of the comprehension variability for individual Ss (Table 1 and Fig. 3C).

Another related measure, the correlation coefficient between the two power spectra (Fcc), could predict about 76% of the variability in sentence comprehension. For comparison, the average power of the MEG signals, measured by the rms of response amplitudes (Table 1 and Fig. 3, magenta curves), could not predict any significant part of this variability.

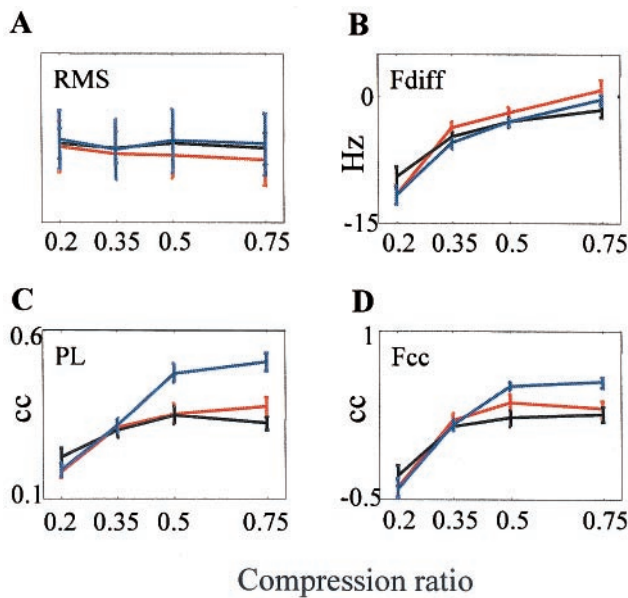
The main predictive power of the stimulus-cortex frequency-matching came from the fact that cortical-evoked modulations, whose frequencies were usually close to the frequency of the envelope at normal speech rates (<10 Hz), remained at the same or lower frequency when the stimulus frequency increased with compression. Thus, with compression, both Fdiff and comprehension were reduced. Although this behavior was consistent across Ss, the actual frequency range that allowed

**Table 1. Potential MEG correlates for speech comprehension**

Correlate	Meaning	Mean, SD, P		
		r	r	value*
rms	Signal power	0.03	0.74	0.90
Fdiff	Stimulus-cortex frequency-matching – difference between modal frequencies	0.94	0.07	0.0000
Fcc	Stimulus-cortex frequency-matching – correlation coefficient between spectra	0.87	0.12	0.0000
PL	Stimulus-cortex phase-locking	0.85	0.16	0.0000

Means and standard deviations of the correlation coefficients between the correlates and comprehension across all Ss, and the probabilities of them reflecting no correlation, are depicted.

\* $P$ [mean( $r$ ) = 0], two-tailed  $t$  test.



**Fig. 4.** Correlates as a function of trial success. Each of the correlates was averaged separately over correct (blue), incorrect (red), and don't know (black) trials across all Ss. Mean  $\pm$  SEM are depicted. Values (rms) are scaled by using arbitrary scaling.

for good comprehension varied among Ss. Interestingly, so did their Fdiffs. The covariation across Ss between Fdiff and *C* is demonstrated in Fig. 3D, in which each symbol represents the threshold values (compression ratio yielding 0.75 of maximal value) of the two variables for an individual S. The linear regression (slope = 0.6,  $r = 0.72$ ,  $P = 0.005$ ) indicates that Fdiff can explain 52% of the variability in *C* across Ss.

The relevance of PL to speech comprehension was examined by determining the cross correlation between the two time-domain signals, i.e., the temporal envelope of the speech input and the temporal envelope of the recorded cortical response (Fig. 2A). The strength of PL was quantified as the peak to peak amplitude of the cross correlation function, filtered at  $\pm 1$  octave around the stimulus modal frequency, within the range 0–0.5 s (Fig. 2C). This measure (PL), which represented the stimulus-cortex time locking at the stimulus frequency band, was also strongly correlated with comprehension (Table 1 and Fig. 3, blue curves). Moreover, the correlation coefficient between *C* and PL was not statistically different from that between *C* and Fdiff ( $P > 0.1$ , two-tailed *t* test).

The low signal to noise ratio of MEG signals did not permit a trial by trial analysis in this study. However, some trial-specific information could be obtained by comparing correct trials vs. incorrect and don't know trials. This comparison revealed that PL was significantly higher during correct than during incorrect trials (2-way ANOVA,  $P = 0.0005$ ) or don't know trials ( $P = 0.0001$ ; Fig. 4), whereas Fdiff was not (2-way ANOVA,  $P > 0.1$ ). The Fcc showed more significant differences than Fdiff, but less significant than PL, between correct, incorrect, and don't know trials (Fig. 4D, 2-way ANOVA,  $P = 0.07$  and 0.01, respectively).

## Discussion

Comprehension of time-compressed (TC) speech was determined by using a variety of speech compression methods (6, 7). These studies have shown that comprehension in normal Ss begins to degrade around a compression of 0.5. However, most previous methods of speech compression did not use compression stronger than 0.4 or 0.3. Here, we used a technique for

speech compression that used a time-scale compression algorithm, which preserved spectral and pitch content across different compression ratios. We were thereby able to compress speech down to 0.1 of its original duration with only negligible distortions of spectral content. This technique allowed us to derive complete psychometric curves, because compressions to 0.2 or less almost always resulted in chance-level performance. In this study, only four compression ratios were used to allow for the averaging of the MEG signals over a sufficient number of trials. Compression ratios were selected so that they spanned the entire range of performance (compressions between 0.2 and 0.75) across all Ss. The psychophysical results obtained were consistent with those obtained in previous TC speech-stimulus studies. However, an additional insight was obtained regarding the neuronal basis of the failures of comprehension for strongly compressed speech.

The main finding was that frequency-matching and PL between the speech envelope and the MEG signal recorded from the auditory cortex were strongly correlated with speech comprehension. This finding was consistent across a group of Ss that exhibited a wide range of reading- and speech-processing abilities. Thus, regardless of the overall performance level, when the comprehension of a given S was degraded because of time compression, so too were the frequency-matching and PL between recorded auditory cortex responses and the temporal envelopes of applied speech stimuli (see Fig. 3). Although both measures gave a good prediction for average comprehension, only stimulus-cortex PL was significantly lower during erroneous trials compared with correct trials. This difference suggests that the capacity for frequency-matching, attributed to the achievable modulation response properties of auditory neurons, is an *a priori* requirement, whereas PL is an “online” requirement for speech comprehension.

A recent study has shown that with sufficiently long stimulus trains, thalamic and cortical circuits can adjust their response frequencies to match different modulation rates of external stimuli (37). However, with short sentences such as those presented here, presumably there was not sufficient time for the brain to change its response frequency according to the stimulus frequency, and it was therefore crucial that the input frequency fall within the effective operational range of *a priori* modulation characteristics of primary auditory cortex neurons. Stimulus-response PL is usually initiated by the first syllable that follows a silent period. Subsequently, if the speech rate closely matches the cortical *a priori* temporal tuning, PL will be high because stimulus and cortical frequencies will correspond. However, if the speech rate is faster than the *a priori* cortical tuning, PL will be degraded or lost (see Fig. 2).

This interpretation is consistent with the successive-signal response characteristics of auditory cortical neurons (e.g., refs. 38 and 39). Interestingly, the strongest response locking to a periodic input is usually achieved for stimulus rates (frequencies) within the dominant range of spontaneous and evoked cortical oscillations, i.e., for frequencies below 14 Hz (40, 41). This is also the frequency range that usually characterizes species-specific calls (42–44). In fact, response locking is strongest for neurons whose best spectral frequency is within the spectrum of the self calls (ref. 42 and S.N., unpublished data). Our results suggest that cortical response locking to the temporal structure of the speech envelope is a prerequisite for speech comprehension. This stimulus-response PL may enable an internal segmentation of different word and sentence components (mostly syllables; see Fig. 1). It is hypothesized that precise PL reflects the segmentation of the sentence into time chunks representing successive syllables, and that in that segmented form spectral analysis is more efficient (45). As mentioned earlier, speech-perception mechanisms have to deal with varying speech rates. Furthermore, different listen-

ers operate successfully within very different ranges of speech rates (Fig. 3D). Our results suggest that for each S, the decodable range is the range of speech rates at which stimulus-cortex temporal locking can be achieved (Figs. 3 and 4).

The neural mechanisms underlying PL and its utilization for speech perception are still not understood. The frequency range of speech envelopes is believed to be too low for the operation of temporal mechanisms based on delay lines (46). However, mechanisms based on synaptic or local circuit dynamics (47) or those based on neuronal periodicity (phase-locked loops; ref. 48) could be appropriate. The advantage of the former mechanisms is that they do not require specialized neuronal elements such as oscillators. The advantage of the latter mechanism is its online adaptivity, which allows tracking of continuous changes in the rate of speech. Recent evidence from the somatosensory system of the rat supports the implementation of phase-locked loops within thalamocortical loops (37, 49–51). Because the computational tasks, and frequency ranges, are similar, similar mechanisms could be useful for both somatosensory and auditory (and maybe even visual; see ref. 52) processing of temporally encoded information.

## Conclusions

We show here that the poor comprehension of accelerated speech is paralleled by a limited capacity of auditory cortical responses to follow the frequency and phase of the temporal envelope of the speech signal. These results suggest that the frequency tuning of cortical neurons determines the upper limit of speech modulation rate that can be followed by an individual, and that a comprehension of a sentence requires an online phase-locking to the temporal envelope of that sentence. Our results, together with recent indications that temporal following is plastic in the adult (53, 54), suggest that training may enhance cortical temporal locking capacities, and consequently, may enhance speech comprehension under otherwise challenging listening conditions.

We thank Susanne Honma and Tim Roberts for technical support, Kensuke Sekihara for providing his software for MUSIC analysis of localization data, and Hagai Attias for his help with data analysis. This work was supported by the Whitaker Foundation (USA), the Dominic Foundation (Israel), and by the U.S.-Israel Binational Science Foundation (Israel).

- Rosen, S. (1992) *Philos. Trans. R. Soc. London B Biol. Sci.* **336**, 367–373.
- Houtgast, T. & Steeneken, H. J. M. (1985) *J. Acoust. Soc. Am.* **77**, 1069–1077.
- Drullman, R., Festen, J. M. & Plomp, R. (1994) *J. Acoust. Soc. Am.* **95**, 1053–1064.
- van der Horst, R., Leeuw, A. R. & Dreschler, W. A. (1999) *J. Acoust. Soc. Am.* **105**, 1801–1809.
- Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J. & Ekelid, M. (1995) *Science* **270**, 303–304.
- Foulke, E. & Sticht, T. G. (1969) *Psychol. Bull.* **72**, 50–62.
- Beasley, D. S., Bratt, G. W. & Rintelmann, W. F. (1980) *J. Speech Hear. Res.* **23**, 722–731.
- Miller, J. L., Grosjean, F. & Lomanto, C. (1984) *Phonetica* **41**, 215–225.
- Dupoux, E. & Green, K. (1997) *J. Exp. Psychol. Hum. Percept. Perform.* **23**, 914–927.
- Newman, R. S. & Sawusch, J. R. (1996) *Percept. Psychophys.* **58**, 540–560.
- Ahissar, M., Protopapas, A., Reid, M. & Merzenich, M. M. (2000) *Proc. Natl. Acad. Sci. USA* **97**, 6832–6837.
- Tallal, P. & Piercy, M. (1973) *Nature (London)* **241**, 468–469.
- Aram, D. M., Ekelman, B. L. & Nation, J. E. (1984) *J. Speech Hear. Res.* **27**, 232–244.
- Shapiro, K. L., Ogden, N. & Lind-Blad, F. (1990) *J. Learn. Disabil.* **23**, 99–107.
- Bishop, D. V. M. (1992) *J. Child Psychol. Psychiatr.* **33**, 2–66.
- Tallal, P., Miller, S. & Fitch, R. H. (1993) *Ann. N.Y. Acad. Sci.* **682**, 27–47.
- Farmer, M. E. & Klein, R. M. (1995) *Psychonomics Bull. Rev.* **2**, 460–493.
- Ben-Yehudah, G., Sackett, E., Malchi-Ginzberg, L. & Ahissar, M. (2001) *Brain* **124**, 1381–1395.
- Watson, M., Stewart, M., Krause, K. & Rastatter, M. (1990) *Percept. Mot. Skills* **71**, 107–114.
- Freeman, B. A. & Beasley, D. S. (1978) *J. Speech Hear. Res.* **21**, 497–506.
- Riensch, L. L. & Clauser, P. S. (1982) *J. Aud. Res.* **22**, 240–248.
- McAnally, K. I., Hansen, P. C., Cornelissen, P. L. & Stein, J. F. (1997) *J. Speech Lang Hear Res.* **40**, 912–924.
- Welsh, L. W., Welsh, J. J., Healy, M. & Cooper, B. (1982) *Ann. Otol. Rhinol. Laryngol.* **91**, 310–315.
- Tiitinen, H., Sivonen, P., Alku, P., Virtanen, J. & Naatanen, R. (1999) *Brain Res. Cogn. Brain Res.* **8**, 355–363.
- Mathiak, K., Hertrich, I., Lutzenberger, W. & Ackermann, H. (1999) *Brain Res. Cognit. Brain Res.* **8**, 251–257.
- Gootjes, L., Raji, T., Salmelin, R. & Hari, R. (1999) *NeuroReport* **10**, 2987–2991.
- Salmelin, R., Schnitzler, A., Parkkonen, L., Biermann, K., Helenius, P., Kiviniemi, K., Kuikka, K., Schmitz, F. & Freund, H. (1999) *Proc. Natl. Acad. Sci. USA* **96**, 10460–10465.
- Joliot, M., Ribary, U. & Llinas, R. (1994) *Proc. Natl. Acad. Sci. USA* **91**, 11748–11751.
- Nagarajan, S., Mahncke, H., Salz, T., Tallal, P., Roberts, T. & Merzenich, M. M. (1999) *Proc. Natl. Acad. Sci. USA* **96**, 6483–6488.
- Patel, A. D. & Balaban, E. (2000) *Nature (London)* **404**, 80–84.
- Woodcock, R. (1987) *Woodcock Reading Mastery Tests—Revised* (Am. Guidance Serv., Circle Pines, MN).
- Portnoff, M. R. (1981) *IEEE Trans. Acoust.* **29**, 1374–1390.
- Mosher, J. C., Lewis, P. S. & Leahy, R. M. (1992) *IEEE Trans. Biomed. Eng.* **39**, 541–557.
- Sekihara, K., Poeppel, D., Marantz, A., Koizumi, H. & Miyashita, Y. (1997) *IEEE Trans. Biomed. Eng.* **44**, 839–847.
- Reite, M., Adams, M., Simon, J., Teale, P., Sheeder, J., Richardson, D. & Grabbe, R. (1994) *Brain Res. Cognit. Brain Res.* **2**, 13–20.
- Pantev, C., Hoke, M., Lehnertz, K., Lutkenhoner, B., Anogianakis, G. & Wittkowski, W. (1988) *Electroencephalogr. Clin. Neurophysiol.* **69**, 160–170.
- Ahissar, E., Sosnik, R. & Haidarliu, S. (2000) *Nature (London)* **406**, 302–306.
- Schreiner, C. E. & Urbas, J. V. (1988) *Hear Res.* **32**, 49–63.
- Eggermont, J. J. (1998) *J. Neurophysiol.* **80**, 2743–2764.
- Ahissar, E. & Vaadia, E. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 8935–8939.
- Cotillon, N., Nafati, M. & Edeline, J.-M. (2001) *Hear Res.* **142**, 113–130.
- Bieser, A. (1998) *Exp. Brain Res.* **122**, 139–148.
- Steinschneider, M., Arezzo, J. & Vaughan, H. G., Jr. (1980) *Brain Res.* **198**, 75–84.
- Wang, X., Merzenich, M. M., Beitel, R. & Schreiner, C. E. (1995) *J. Neurophysiol.* **74**, 2685–2706.
- van den Brink, W. A. & Houtgast, T. (1990) *J. Acoust. Soc. Am.* **87**, 284–291.
- Carr, C. E. (1993) *Annu. Rev. Neurosci.* **16**, 223–243.
- Buonomano, D. V. & Merzenich, M. M. (1995) *Science* **267**, 1028–1030.
- Ahissar, E. (1998) *Neural Comput.* **10**, 597–650.
- Ahissar, E., Haidarliu, S. & Zacksenhouse, M. (1997) *Proc. Natl. Acad. Sci. USA* **94**, 11633–11638.
- Sosnik, R., Haidarliu, S. & Ahissar, E. (2001) *J. Neurophysiol.* **86**, 339–353.
- Ahissar, E., Sosnik, R., Bagdasarian, K. & Haidarliu, S. (2001) *J. Neurophysiol.* **86**, 354–367.
- Ahissar, E. & Arieli, A. (2001) *Neuron* **32**, 185–201.
- Kilgard, M. P. & Merzenich, M. M. (1998) *Nat. Neurosci.* **1**, 727–731.
- Shulz, D. E., Sosnik, R., Ego, V., Haidarliu, S. & Ahissar, E. (2000) *Nature (London)* **403**, 549–553.