

# Speech Modifications Algorithms Used for Training Language Learning-Impaired Children

Srikantan S. Nagarajan, Xiaoqin Wang, *Member, IEEE*, Michael M. Merzenich, Christoph E. Schreiner, Paul Johnston, William M. Jenkins, Steven Miller, and Paula Tallal

**Abstract**—In this paper, the details of processing algorithms used in a training program with language learning-impaired children (LLI's) are described. The training program utilized computer games, speech/language training exercises, books-on-tape and educational CD-ROM's. Speech tracks in these materials were processed using these algorithms. During a four week training period, recognition of both processed and normal speech in these children continually increased to near age-appropriate levels. We conclude that this form of processed speech is subject to profound perceptual learning effects and exhibits widespread generalization to normal speech. This form of learning and generalization contributes to the rehabilitation of temporal processing deficits and language comprehension in this subject population.

**Index Terms**—Acoustic processing, dyslexia, language learning, modulation processing, perceptual learning, specific language impairments, speech processing, temporal processing.

## I. INTRODUCTION

LANGUAGE learning-impaired children (LLI's) have a specific phonological processing deficit marked by difficulty in recognizing the phonetic elements of speech delivered in fast natural input sequence forms. It has been hypothesized that some LLI's have a rapid processing deficit that results in their limited ability to recognize and process short and rapidly occurring acoustic elements in specific acoustic contents in running speech [1]. For example, many LLI's have adequate time-order judgment, backward masking and modulated input-fusion capabilities for relatively long duration stimulus events presented in sequence. However, they break down in their stimulus segmentation and stimulus event recognition at higher input rates and with shorter-duration events [2], [3]. LLI children commonly cannot identify fast elements in speech that have durations in the range from 10 to 50 ms, a critical time frame over which many phonetic contrasts are embedded. For example, they have difficulty in discriminating between

speech syllables such as /ba/ and /da/ in which the initial consonants are characterized by rapid frequency transitions that occur within a few tens of milliseconds, as well as /da/ and /ta/ in which the consonants are cued by voice onset time.

Based on research on cortical plasticity, we hypothesized that learning disabilities arise by normal learning mechanisms operating on either defective inputs or operating aberrantly on normal inputs [4]. Given an experiential history that has hypothetically strongly embedded defective spatio-temporal processing widely across the forebrain, how can we hope to reverse such defective processing? As a first approach, we developed adaptive training exercises, applied in the form of CD-ROM mounted games designed to drive progressive improvements in LLI children's temporal processing abilities [5]. In parallel, we also provided LLI children with processed speech inputs that therefore 1) were more salient and more discriminable for this subject population and 2) modified to provide more sharply temporally disambiguated inputs, which we hypothesized would provide stronger learning inputs for creating a more robust phonetic element representation within the cortical learning machinery [6]. We reasoned that sharpening the waveform modulation structure should provide a basis for learning a sharper segmentation of successive phonetic elements [4].

In this paper, we describe in detail, the algorithms used to create this form of modified speech. Additionally, we report gradual improvements in processed speech comprehension in LLI's. The profound effects of learning such modified speech form taken together with earlier reports about improvements in normal speech and language comprehension suggests an useful and novel application for perceptual learning of complex spatio-temporal stimuli in rehabilitation [6].

## II. THEORY

We used a two-stage speech modification procedure. The first stage involved time-scale modification of speech signals without altering their spectral content. We implemented a commonly used time-scaling algorithm called the "phase vocoder" that will be reviewed briefly in Section I. Prolongation of the speech stream, especially the formant transitions occurring in consonants has been shown to increase intelligibility for LLI's [2]. However, prolonging speech alone does not render faster elements in speech more salient and they can still be subject to forward or backward masking by neighboring slowly modulated speech elements. Moreover, primate studies have revealed that engaging the brain in making distinctions

Manuscript received July 3, 1997; revised February 12, 1998 and May 14, 1998. This work was supported by a grant from the Charles Dana Foundation and by the Hearing Research, Inc.

S. S. Nagarajan, M. M. Merzenich, and W. M. Jenkins are with the Keck Center for Integrative Neuroscience, University of California at San Francisco, San Francisco, CA USA. They are also with the Scientific Learning Corporation, Berkeley, CA USA.

X. Wang is with the Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD USA.

C. E. Schreiner and P. Johnston are with the Keck Center for Integrative Neuroscience, University of California at San Francisco, San Francisco, CA USA.

S. Miller and P. Tallal are with the Scientific Learning Corporation, Berkeley, CA USA. They are also with Rutgers University, Newark, NJ USA.

Publisher Item Identifier S 1063-6528(98)05924-2.

about slowly modulated inputs on a heavy schedule results in changes that favor slow modulation-rate processing and disfavours rapid modulation-rate (e.g., normal fast speech) processing [7], [8]. Thus, exposure to only prolonged speech signals by LLI's would be counter-productive for learning a more reliably segmented representation of successive, fast phonetic elements. Therefore we devised an algorithm that differentially amplifies and disambiguates faster phonetic elements in speech. Fast elements in speech were defined as those changes that occurred in the 3–30 Hz range within the envelope of narrow-band “channels” of the rate-changed speech signal. An “emphasis” algorithm for selective amplification of these fast elements was implemented using two methods: a filter-bank summation method and an overlap-add method based on the short-time Fourier transform. We describe these two algorithms for differential emphasis of fast-elements in speech in Sections II and III.

### III. SECTION I:

#### TIME-SCALE MODIFICATION ALGORITHM

Time-scale modification of speech signals was accomplished using the phase-vocoder algorithm originally proposed by Flanagan [9] with some of the modifications suggested by Portnoff using the short-time discrete Fourier transform [10]–[12]. A broadband segment of speech assumed to be composed of a set of narrow-band signals each obtained by passing the speech segment through a filter bank of band-pass filters. Thus, we can write the speech signal  $f(t)$  as follows:

$$f(t) \cong \sum_{k=1}^N f_k(t) \quad (1)$$

where

$$f_k(t) = \int_{-\infty}^t f(t)h(t-\tau) \cos[\omega_k(t-\tau)] d\tau \quad (2)$$

This is the convolution integral of the signal  $f(t)$  and  $h(t)$ , a prototypical low-pass filter modulated by  $\cos[\omega_k(t)]$  where  $\omega_k$  is the center frequency of the filters in the filter-bank (this transformation is commonly referred to as heterodyning). This integral can be calculated from the windowed short-time Fourier transform of the input signal evaluated at the radian frequency  $\omega_n$  using the fast Fourier transform (FFT) algorithm [10], [13]–[15]. If we denote the complex value of this transform as  $(\omega_k, t)$ , then

$$f_k(t) = |F(\omega_k t)| \cos[\omega_k t + \varphi_k(\omega_k, t)] \quad (3)$$

where  $\varphi_k(\omega_k, t)$  is the phase modulation of the carrier  $\cos[\omega_k(t)]$ . It has been shown that the phase function is not a well behaved function, but its derivative, the instantaneous frequency, is bounded and is band-limited [16]. Therefore, a practical approximation for  $f_k(t)$  is

$$f_k(t) \cong |F(\omega_k, t)| \cos \left[ \omega_k t + \int_0^t \dot{\varphi}_k(\omega_k, \tau) d\tau \right] \quad (4)$$

where  $\varphi$  is the instantaneous frequency and can be computed from the unwrapped-phase of the short-time Fourier transform.

A time-scaled signal can then be synthesized as follows by interpolating the short-time Fourier transform magnitude and the unwrapped phase to the new time scale

$$f(\beta t) \cong \sum_{k=0}^K |F(\omega_k, \beta t)| \cos \left( \beta \omega_k t + \int_0^t \dot{\varphi}_k(\omega_k, \beta \tau) d\tau \right). \quad (5)$$

where  $\beta$  is the scaling factor which is greater than one for time-scale expansion. An efficient method to solve (5) is to use cyclic rotation and the FFT algorithm along with an overlap-add procedure to compute the short-time discrete Fourier transform [11], [14], [17]. The choice of the analysis filters and interpolating filters (for interpolation of the short-time Fourier-transformed data to the new time-scale) affect the sensitivity and performance of the algorithm. For simplicity, instead of using interpolation based on a synthesis filter as originally proposed by Portnoff, we used linear interpolation on the magnitude and the phase of the short-time Fourier transform as described by Gordon *et al.* [18]. The analysis filter  $h(t)$  was chosen to be a Kaiser window multiplied by an ideal impulse response [15].

This time-scale modification algorithm implemented on a Silicon Graphics Indy workstation, was slower than Portnoff's algorithm because of the linear-interpolation synthesis and performed at 10–15 slower than real-time for time-stretching by 50%. Example speech waveforms tested with this algorithm are shown in Fig. 1. Time-scaling factors used in the training program were varied from 1.5 times normal at the beginning of the program to 1.2 toward the end of the program, with an FFT size of 256 and 512 samples for audio tracks sampled at frequencies of 22.05 and 44.1 kHz, respectively. For 22.05 kHz tracks, the FFT sizes were increased to 512 when we used male voices with very low pitch to reduce extraneous reverberations. The length of the Kaiser window was set at 9 ms. Informal listening tests performed on normal adults indicated that, consistent with the literature, this algorithm performed with high fidelity and did not result in any perceptual loss in speech intelligibility. Characteristics such as speaker identity are well preserved by the processing.

### IV. SECTION II: FILTER-BANK EMPHASIS ALGORITHM

As a first attempt at differential amplification of these fast modulation envelopes we implemented a digital signal processing algorithm using a filter-bank filtering method. Again, we assumed that the speech signal can be synthesized from a set of narrow-band signals which were obtained by passing the original signal through a bank of band-pass filters as given in (1). This time however, we did not use heterodyning of a prototypical low-pass filter. Instead, we used a set of up to 20 second-order Butterworth filters with center frequencies logarithmically spaced between 100 Hz and the Nyquist frequency as shown in Fig. 2(a). The output of each band-pass filter resulted in a narrow-band channel signal is given by

$$f_k(n) = f(n) * h_k(n) \quad (6)$$

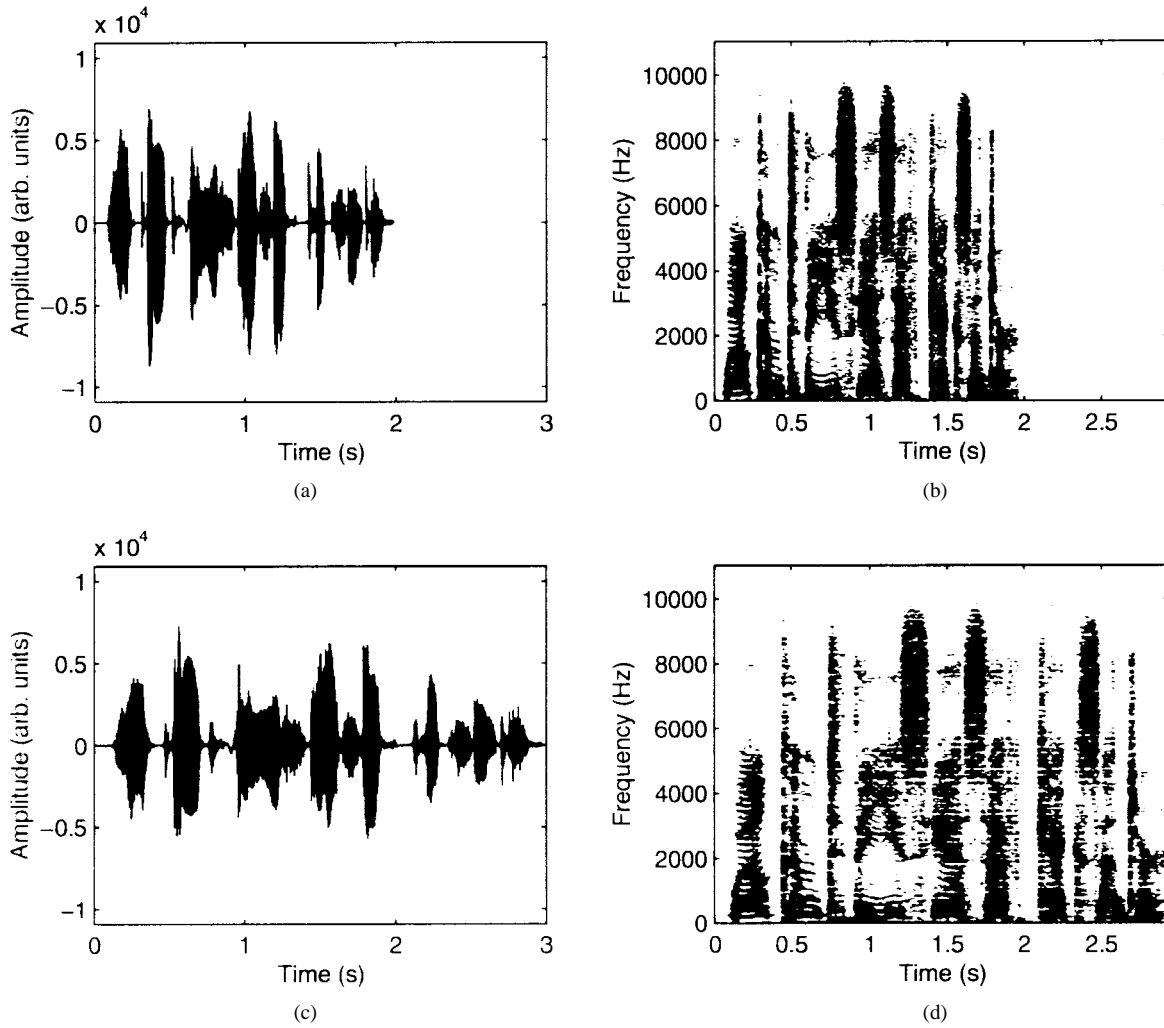


Fig. 1. Exemplary waveforms and spectrograms used for the time-scale modification algorithm. (a) and (b) Waveform and spectrograms for the sentence “Welcome to the UCSF Keck Center” sampled at 22.05 kHz. (c) and (d) Waveform and spectrograms for the same sentence stretched by 1.5 times using an analysis window of 256 samples and an overlap window of 128 samples.

where  $h_k(n)$  is the corresponding band-pass filter and  $*$  indicates the convolution operator. We then computed the analytical signal as follows:

$$a_k(n) = f_k(n) + i\mathbf{H}(f_k(n)) \quad (7)$$

where  $\mathbf{H}(n)$  is the Hilbert transform of a signal defined as

$$\mathbf{H}(n) = f_k(n) * \left( \frac{1}{\pi t} \right) = \int f_k(\tau) \frac{1}{\pi(n - \tau)} d\tau. \quad (8)$$

The Hilbert transform was computed using the FFT algorithm. It has been shown that the absolute value of the analytical signal is the envelope of a narrow-band signal [15]. Thus, we obtained the envelope  $e_k(n)$  within each narrow-band given by the absolute value of the analytical signal ( $e_k(n) = |a_k(n)|$ ). The envelope within each narrow-band channel was then band-pass filtered using a second-order Butterworth filter with cutoffs set between 3–30 Hz. Forward and backward filtering operations were performed to minimize phase distortions. The band-pass filtered envelope was then rectified to form the new envelope as follows:

$$e_k^{\text{new}}(n) = S(e_k(n) * g(n)) \quad (9)$$

where  $*$  is the convolution operator and  $S$ , is the rectification function

$$S(x) = \begin{cases} x & x \geq 0 \\ 0 & x < 0 \end{cases} \quad (10)$$

and  $g(n)$  is the impulse response of the band-pass second order Butterworth filter. We then modified the signal within each band-pass channel to carry this new envelope

$$s_k^{\text{new}}(n) = \left[ f_k(n) S \left( \frac{e_k^{\text{new}}(n)}{e_k(n)} \right) \right]. \quad (11)$$

To prevent distortions due to the rectification from frequencies outside the pass-band, this modified signal was band-pass filtered with a second-order Butterworth filter and normalized to carry the same power within the band-pass region

$$f_k^{\text{new}}(n) = \frac{\|f_k(n)\|}{\|s_k^{\text{new}}(n)\|} s_k^{\text{new}}(n) * h_k(n). \quad (12)$$

The final modified signal was then obtained by summing the narrow-band filters with a differential gain for each channel

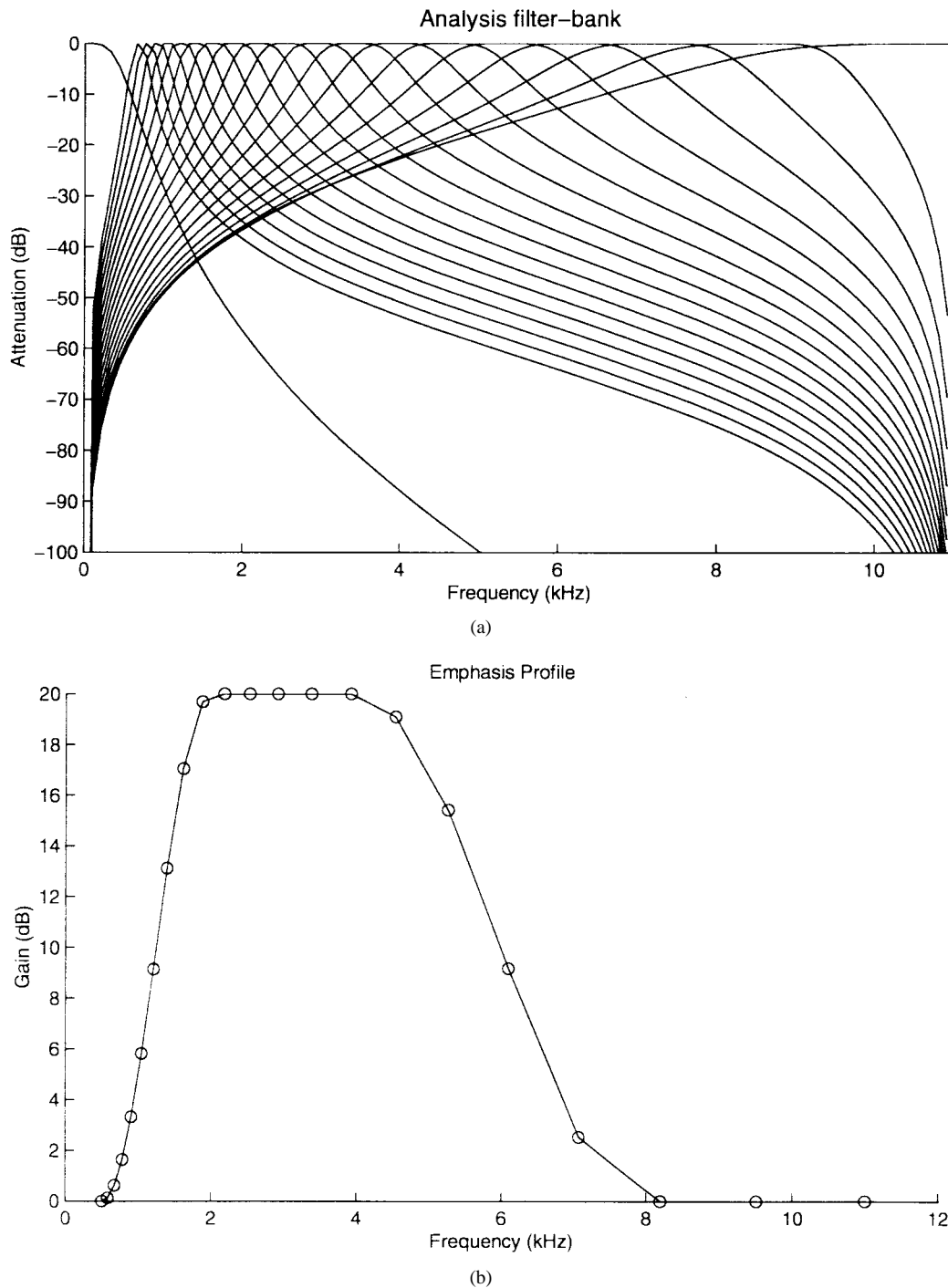


Fig. 2. (a) Frequency response of filters used in the filter-bank. Second-order Butterworth filters with center-frequencies logarithmically separated between 1–10 kHz were used. (b) The profile of the gain for each channel in the filter-bank. Note a maximum of 20 dB gain for the channels in the second-formant range.

as follows:

$$f_k^{\text{new}}(n) = \sum_k w_k f_k^{\text{new}}(n) \quad (13)$$

where  $w_k$  is the added gain for each channel as shown in Fig. 2(b). Additionally, to enhance the speech signals in the second formant range important for speech intelligibility, a maximum gain of 20 dB was imposed in the range of the

second formant between 1–4 kHz. An example of the final result of processing by this algorithm is shown in Fig. 3.

This algorithm was also implemented in a Silicon Graphics Indy workstation and performed at a speed from 40 to 50 times slower than real-time. Again, informal listening tests were conducted on normal adults. The algorithm resulted in a synthetic-sounding speech output and a shift in timbre due to the amplification of the second-formant frequencies. However, there was no reported loss in speech intelligibility in normal adults with this form of modified speech.

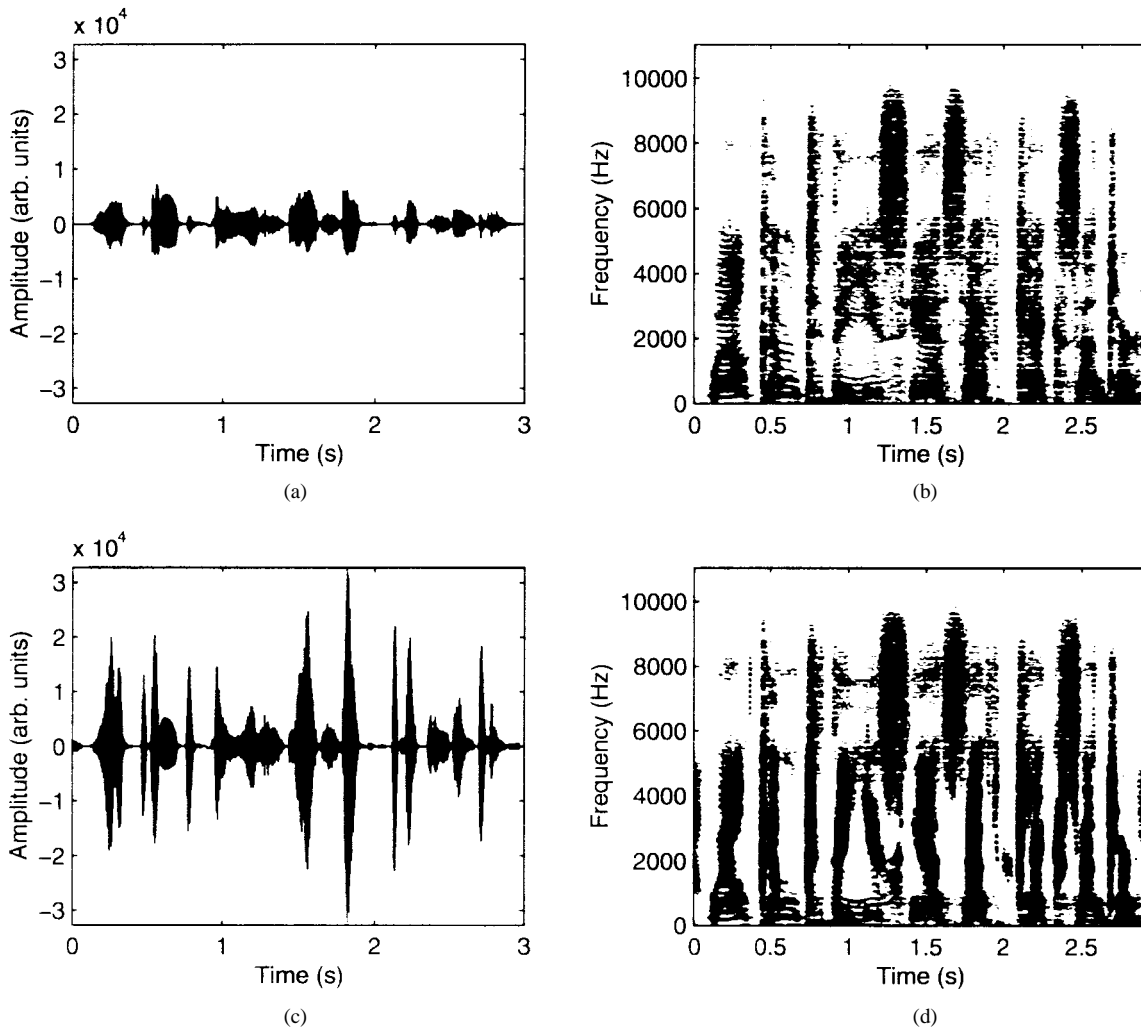


Fig. 3. Example waveforms and spectrograms used by the filter-bank emphasis algorithm. (a) and (b) Waveform and spectrograms for the sentence “Welcome to the UCSF Keck Center” prolonged by 1.5 times and sampled at 22.05 kHz. (c) and (d) Waveform and spectrograms for the same sentence processed by the filter-bank emphasis algorithm.

### V. SECTION III: OVERLAP-ADD EMPHASIS ALGORITHM

We improved the speed of the filter-bank summation by making use of the property of equivalence between the short-time Fourier transform and the filter-bank summation algorithm [13], [17]. In this case, the short-time Fourier transform was computed using an overlap-add procedure and the FFT algorithm [14]. The short-time Fourier transform computed over a sliding window is given by the following equation:

$$X_k(r) = \sum_{n=-\infty}^{\infty} h(r-n)x(n)e^{-i2\pi nk/K} \quad (14)$$

where  $h(n)$  is a Hamming window and the overlap between sections was chosen to be less than a quarter the length of the analysis window [13], [17]. We could then directly obtain the envelope within a narrow-band channel from the absolute value of the short-time Fourier transform [16]. The number of narrow-band channels was equal to half the size of the length over which the FFT was computed. We then averaged

the energy of the envelope within critical bands channels

$$f_k(r) = \sum_{C_{k-1} \leq k \leq C_k} |X_k(r)| \quad (15)$$

where  $C_k$  is the corner-frequency of the critical-band channel  $k$ . As a first approximation, we used the parameters for auditory critical bands as proposed by Zwicker [19], [20]. The envelope within each critical-band channel was then band-pass filtered with cutoffs set usually between 3–30 Hz (the time scale at which phonetic events occur in rate-changed speech) with type I linear-phase FIR equiripple filters [15]. This avoided any phase distortions caused in the initial and final segments of the data. The amplitude and phase of the frequency response of this filter are shown in Fig. 4(a) and (b). The band-pass filtered envelope was then threshold rectified. In contrast to the filter-bank emphasis algorithm, the modified envelope was then added to the original envelope to amplify the fast-elements while retaining the slower modulations in their original forms. This is given by the following equation:

$$X_k^{\text{filt}}(n) = X_k(n)T\left(\frac{e_k^{\text{new}}(n)}{e_k(n)}\right) \quad (16)$$

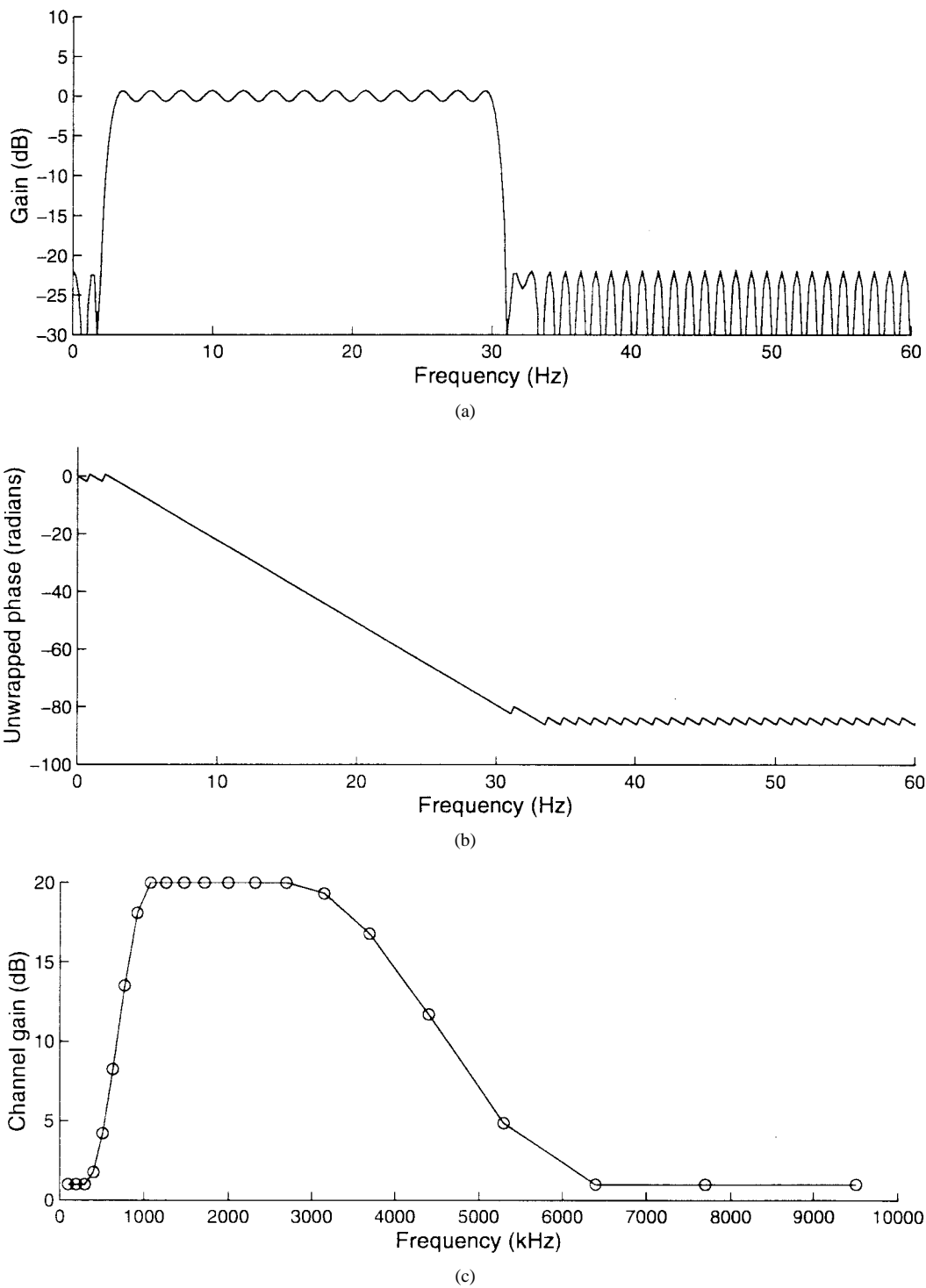


Fig. 4. (a) Magnitude and (b) phase response of the envelope filter between 3–30 Hz designed using Parks-McCelland algorithm. (c) Channel gain for each critical-band channel. Note a maximum of 20 dB gain for the channels in the second-formant range.

where  $T$  is a thresholding function

$$T = \begin{cases} x + 1, & x \geq \theta \\ 1, & x < \theta. \end{cases} \quad (17)$$

The threshold parameter  $\theta$  was chosen to be 80 dB down from the peak envelope amplitude within a frequency channel. Additionally, the signal power within each channel was normalized to have the same power as the original signal within

the narrow-band channel

$$X_k^{new}(n) = \frac{\|X_k(n)\|}{\|X_k^{filt}(n)\|} X_k^{filt}(n). \quad (18)$$

A modified signal was then obtained by summing the short-time Fourier transforms using a weighted overlap-add proce-

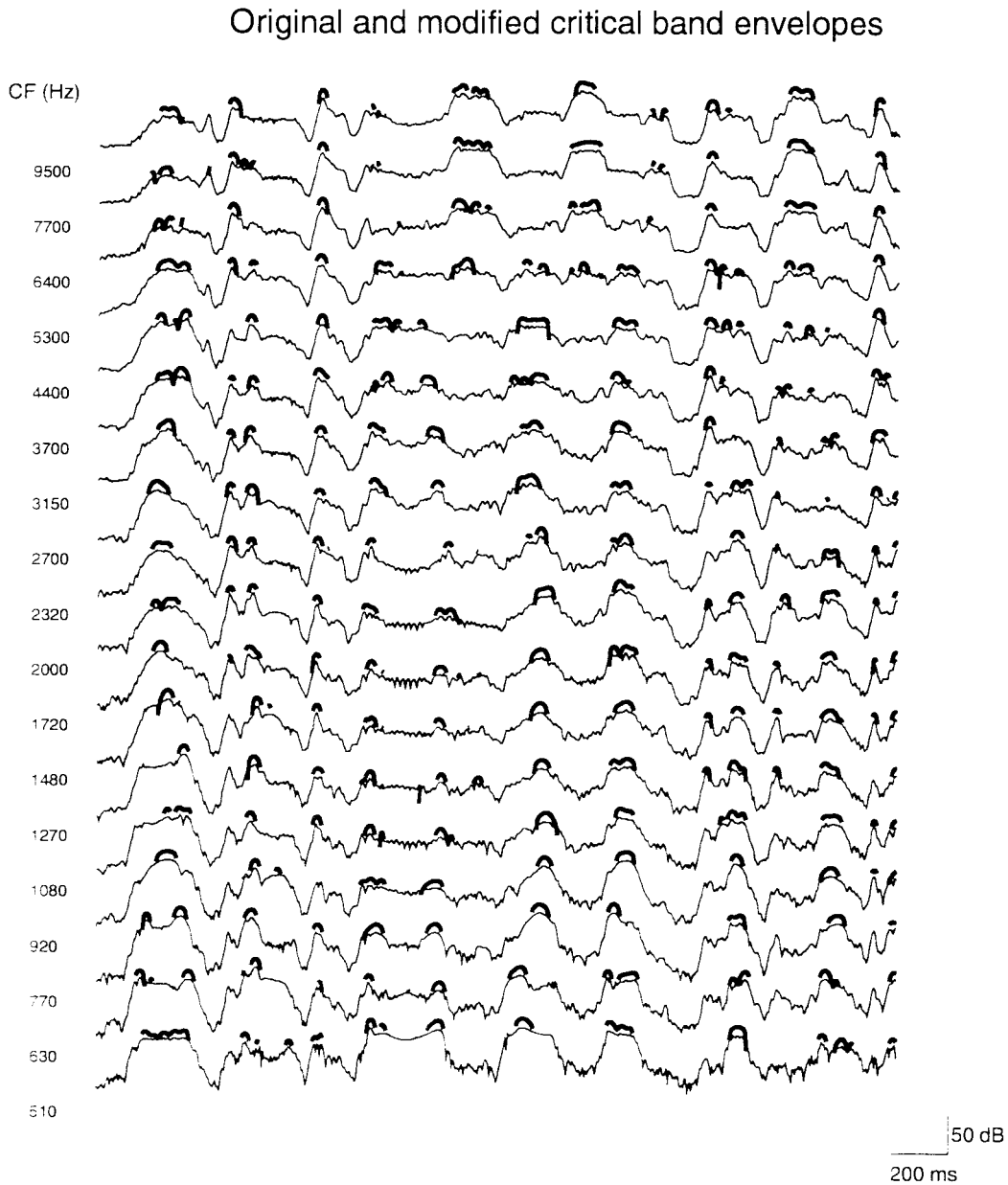


Fig. 5. Original (thin line) envelopes, of the test sentence, within each critical-band channel and the modified segments of each of these envelope due to the filtering and rectification (thick lines). The center-frequency of each critical-band filter is indicated at the left of each envelope.

dure similar to that used for analysis as follows:

$$f^{\text{new}}(n) = \sum_{m=-\infty}^{\infty} g(n-m) \sum_{k=0}^{K-1} X_k^{\text{new}}(m) e^{i2\pi mk/K} \quad (19)$$

where  $g(n)$  is the synthesis filter which was also chosen to be a Hamming window [14]. A typical profile for the gain for each critical-band channel is shown in Fig. 4(c), with a maximum gain of 20 dB within the range of the second formant. An example of the modified segments of the speech envelope is shown in Fig. 5. In this form it is observed that in almost all channels, the changes are selectively made to the faster modulations, and the slower modulations remain mostly unmodified. An example of a processed speech output from this algorithm is shown in Fig. 6.

This algorithm was also implemented in a Silicon Graphics Indy workstation and dramatic improvements were observed in computation time. This algorithm performed at speeds of 1.2 slower than real-time. The output of this algorithm was similar to that of the filter-bank version. Informal listening tests conducted on normal adults indicated that the speech output was synthetic and shifted in the timbre due to the amplification of the second-formant frequencies but resulted in no perceptual loss in speech intelligibility.

## VI. METHODS AND RESULTS

The above-mentioned two-step processing algorithms were applied to the speech in language listening exercises recorded on audiotapes, as well as to the speech tracks of several children's stories recorded on tapes and on educational CD-

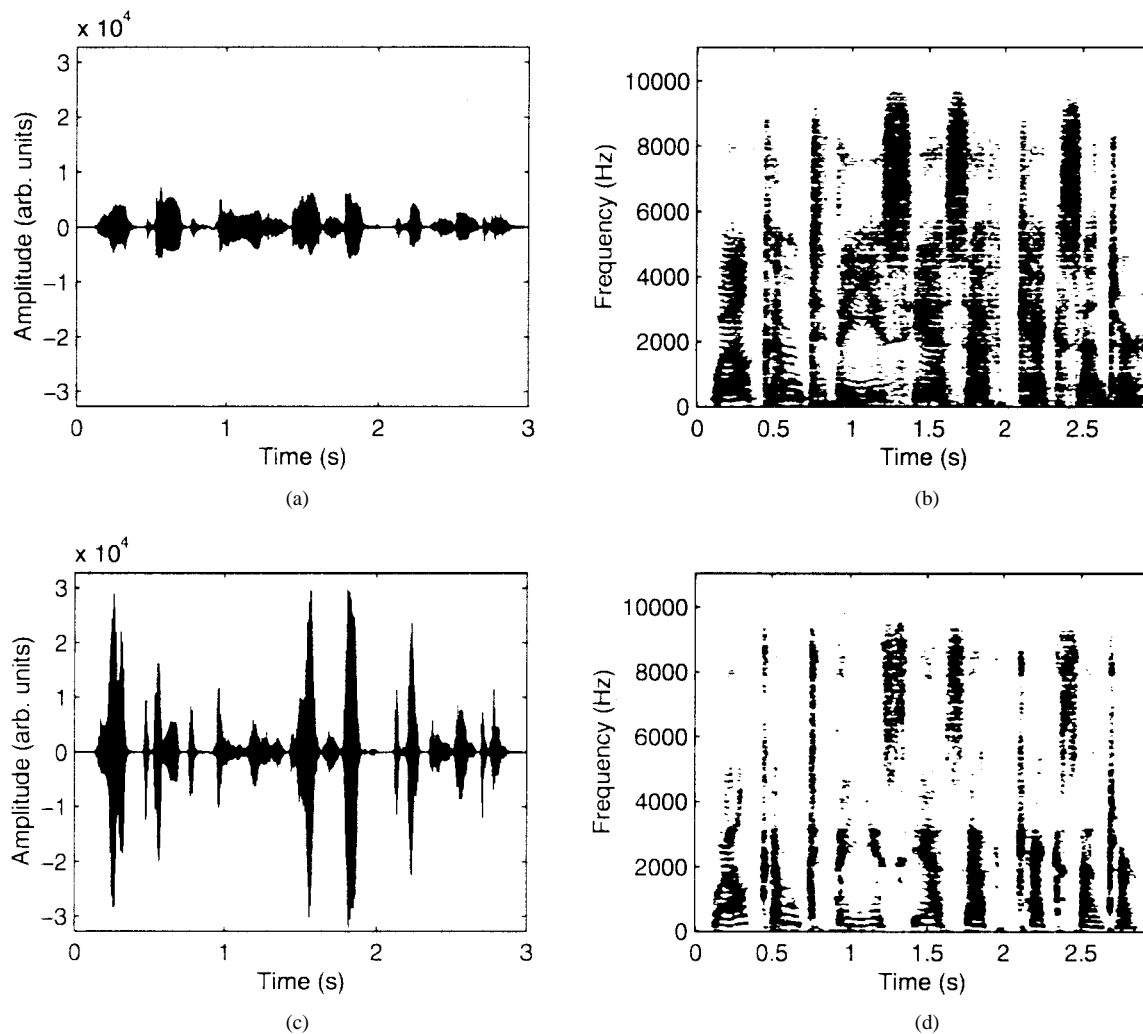


Fig. 6. Example waveforms and spectrograms used by the overlap-add emphasis algorithm. (a) and (b) Waveform and spectrograms for the sentence "Welcome to the University of California at San Francisco (UCSF) Keck Center" prolonged by 1.5 times and sampled at 22.05 kHz. (c) and (d) Waveform and spectrograms for the same sentence emphasized processed by the overlap-add emphasis algorithm.

ROM's. These modified speech stimuli were used extensively in two four-week long studies organized for LLI children [6]. In the first study, seven LLI children (four males and three females) ranging in ages from 5.8 to 9.1 years (mean age =  $7.3 \pm 1.5$ ) participated in a six-week study aimed at evaluating the effect that exposure to this form of modified speech would have on speech discrimination and language comprehension. The children had previously demonstrated severe delays in receptive and expressive language development as well as marked temporal processing deficits. These children also had reading deficits but were without other primary deficits. This group was exposed to modified speech which was time-scale expanded and emphasized by the filter-bank emphasis algorithm.

In the second study, eleven children participated who ranged in age from 5.2 to 10.0 years (mean age =  $7.4 \pm 1.4$ ) and who had a mean nonverbal intelligence score of  $96.4 \pm 9.7$ . Again, all children demonstrated a severe delay in receptive and expressive language development as well as marked temporal deficits and reading deficits but no other primary deficits. This group was exposed to modified speech which was time-

scale expanded and emphasized by the overlap-add emphasis algorithms.

In both the studies, additional computer games were designed with speech and nonspeech stimuli to drive improvements in the impaired child's reception of fast successive acoustic events [5]. During the training period, children spent 3 h/day for 20 days playing these computer games. Training also included practice in making phonetic distinctions, in hierarchical instructional tasks patterned after the Token test for children [6]. Children were also trained in a memory-for-sentences task with sentences of increasing length and linguistic complexity. Finally, children were trained in a hierarchical language game based on the Curtiss-Yamada Comprehensive Language Evaluation (CYCLE). All of this training was administered with acoustically modified speech as described above.

In addition to pre- and post-training improvements reported elsewhere [6], LLI children also showed gradual improvements in their performance on speech and language tests administered during training. These results are shown in Fig. 7. In the first study, the Clinical Evaluation of Spoken Language

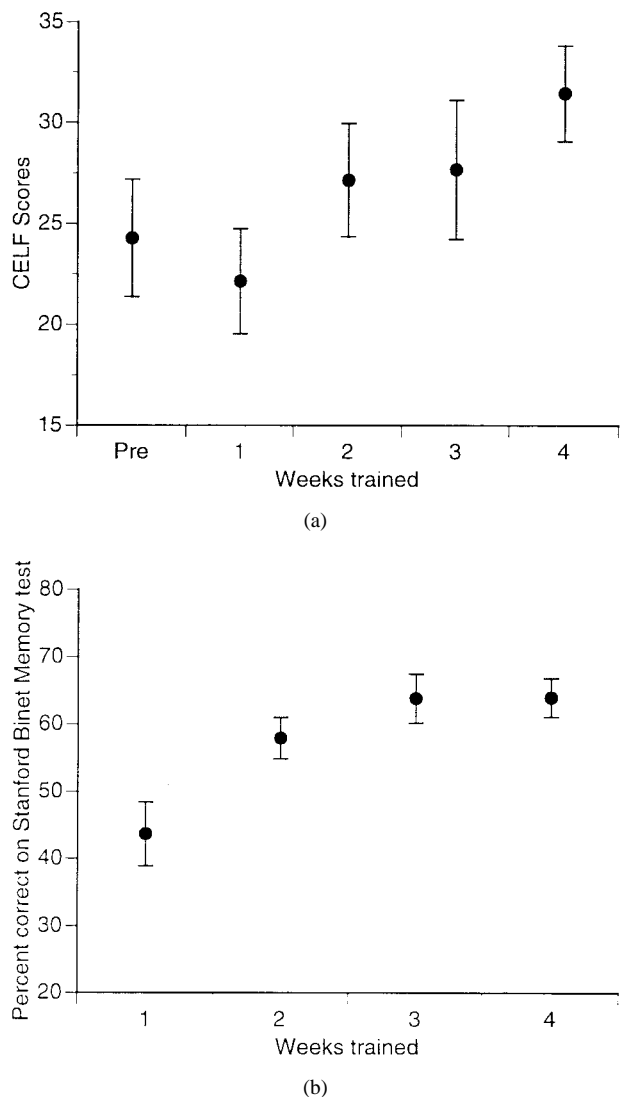


Fig. 7. Learning curves for speech comprehension tests measured weekly during training. (a) For subjects from study 1, mean CELF-R scores. (b) For study 2, mean percent correct scores on a test patterned after Stanford Binet memory for sentences. Error bars indicate one standard error of mean across subjects.

Function (CELF) test was administered at the end of every week of training and the performance on this test showed gradual improvement over the training period [see Fig. 7(a)]. Also shown here are the pretest scores. The data indicates a nonsignificant drop in the test score in the first week of training followed by a gradual and continuous increase in scores in the following weeks. A repeated-measures ANOVA with testing time (including pretest score at week 0) as a repeated measure and threshold as the factor indicated that these improvements were statistically significant ( $F(5, 4) = 6.4, p < 0.001$ ).

In the second study, Stanford-Binet “Memory for Sentences” subtests were administered weekly during the training period. The performance on this test (measured conservatively as the percent correct of number of the sentences presented) showed gradual improvement over the training period [see Fig. 7(b)]. Again, a one-way ANOVA with testing-time as a repeated measure and threshold as a factor indicates that

these improvements were statistically significant ( $F(9, 3) = 15.5, p < 0.001$ ).

## VII. DISCUSSION

The current paper presents a novel application for speech modification algorithms that manipulate the speech envelope to drive central nervous system representational changes of phonetic elements in language-learning impaired children. These results are consistent with other forms of perceptual learning observed in normal adult humans following listening training with phonetic contrasts, where speech discrimination contrasts used for training were manually constructed and practiced by the subjects [21]–[24].

As a result of training, the LLI children significantly improved in their ability to identify rapidly successive sounds and fast, natural consonant-vowel stimuli [5]. These children also improved substantially in their phonetic reception abilities and in language comprehension, as measured by alternative test versions of Glodman-Fristow-Woodcock (GFW) test of auditory comprehension, the Token test and CYCLE [6]. Pre- and post-tests were presented with natural speech and included none of the same items, distinctions or instructions that had been previously presented during training to the children. Children trained with temporally-modified speech showed far greater language gains than did carefully matched control children, who received identical training, but with natural, unmodified speech [6]. These results suggest that the generalization of perceptual learning is widespread. Such widespread generalization is consistent with recent studies on perceptual learning of temporal information in both auditory and somatosensory systems [25], [26].

Several speech enhancement algorithms have been reported in the literature for various applications involving analysis of speech perception [19], [27]–[33]. Our algorithms for speech emphasis are similar to those derived by Langhans for speech enhancement [19]. Langhans *et al.* have shown that similar processing of speech signals, without any time-scale expansion, did not deteriorate speech intelligibility in normal adult subjects. They did not address the issue of driving representational changes by training with this form of modified speech. Our observed increase in speech comprehension performance during training suggests that commonly used measures of speech intelligibility are subject to learning effects. Moreover, our learning results could not be due to adaptation to expanded speech because such adaptation effects have been reported to be short (in the order of minutes) and would not account for the generalization to normal unaltered speech [34]. These effects of learning and generalization have often been ignored in previous studies on evaluation of speech enhancement algorithms.

Our informal listening tests indicate that three components of the modification algorithms—time-scale expansion, envelope filtering and the channel profile gain—do not affect speech intelligibility. These results are consistent with previous reports on the effect of speech intelligibility with comparable modifications. First, reports in normal adults clearly indicate that no loss in intelligibility is observed with time-compressed

and time-expanded speech with compression factors from 50 to 200% [10], [35]. Second, the effects on speech intelligibility of modifications to the speech envelope within narrow-band channels have been extensively studied [36]–[39]. It has been shown that both low-pass and high-pass filtering of the envelope does not cause any deterioration in the intelligibility of speech for frequencies between 4–60 Hz [36]–[39]. Moreover, envelope energy within this frequency range does not significantly affect speech intelligibility even when synthesis is obtained with a noise carrier. Third, enhancement of spectral contrast in speech show moderate gains in recognition of some stop consonants by hearing impaired listeners [28]. The channel gain profile function achieves this effect by selective amplification of frequencies in the second formant range which have been shown to be important for speech intelligibility [40].

The envelope of an acoustical signal has been shown to give rise to cues which are more robust against reverberation and additional noise than the fine structure of the signal [38], [39]. The formation and evaluation of phonetic elements might be based on the modulation structure of the speech envelope [4], [5]. Furthermore, cortical representation of time-varying broadband signals appears to be organized as an abstraction of the envelope spectra in different center-frequency regions [41]. Plasticity of the cortical representation of sounds has primarily been studied with simpler stimuli. Experiments in animals have shown that perceptual learning of frequency discrimination and amplitude modulation results in observable plastic changes in the primary cortical representation of these trained stimuli. These experiments indicate that more sharply modulated speech signals should be more powerful for inducing complex learning [4]. However, further studies are necessary to document the neurophysiological changes resulting from behavioral improvements due to training with complex spatio-temporal stimuli such as speech.

Further evaluation is necessary to determine the relative contributions to learning of the magnitude of time-scaling and the differential emphasis. The computational speed of these algorithms can be increased by combining the time-scale and emphasis algorithms into a single step. Such a change might result in a real-time processing algorithm, potentially leading to a feasible prosthetic device for rehabilitation of this population [42]. However, as suggested by Clarkson and Bahgat [29], the simplifications imposed on a real-time device could potentially counterbalance the effectiveness of nonreal time algorithms. Therefore, further studies are necessary to determine the effects of simplifications that can be achieved for real-time implementation.

In conclusion, training LLI children with speech stimuli in which rapidly changing components have been temporally prolonged and differentially amplified resulted in a dramatic improvement in their speech and language perception abilities. Such form of modified speech appear to provide a powerful input for the rehabilitation and remediation of LLI.

#### ACKNOWLEDGMENT

The authors would like to thank C. Wen, Z. Wu, and S. Wang for assistance in implementation of some

of the algorithms and Dr. A. Protopapas for comments on an earlier version of this manuscript. For additional information on this research, see <http://www.ld.ucsf.edu> and <http://www.scilearn.com>.

#### REFERENCES

- [1] P. Tallal and M. Piercy, "Developmental aphasia: Rate of auditory processing and selective impairment of consonant perception," *Neuropsychol.*, vol. 12, pp. 83–93, 1974.
- [2] P. Tallal and M. Piercy, "Defects of nonverbal auditory perception in children with developmental aphasia," *Nature*, vol. 241, pp. 468–469, 1973.
- [3] B. A. Wright, L. J. Lombardini, W. M. King, C. S. Puranik, C. M. Leonard, and M. M. Merzenich, "Deficits in auditory temporal and spectral processing in language-impaired children," *Nature*, vol. 387, pp. 176–178, 1997.
- [4] M. M. Merzenich, C. Schreiner, W. M. Jenkins, and X. Wang, "Neural mechanisms underlying temporal integration, segmentation, and input sequence representation: Some implications for the origin of learning disabilities," *Ann. New York Academy Sci.*, vol. 682, pp. 1–22, 1993.
- [5] M. Merzenich, W. Jenkins, P. Johnston, C. Schreiner, S. Miller, and P. Tallal, "Temporal processing deficits of language-learning impaired children ameliorated by training," *Sci.*, vol. 271, pp. 77–81, 1996.
- [6] P. Tallal, S. L. Miller, G. Bedi, G. Byma, X. Wang, S. S. Nagarajan, C. Schreiner, W. M. Jenkins, and M. M. Merzenich, "Language comprehension in language-learning impaired children improved with acoustically modified speech," *Sci.*, vol. 271, pp. 81–83, 1996.
- [7] R. E. Beitel, C. E. Schreiner, X. Wang, S. W. Cheung, and M. M. Merzenich, "Amplitude modulated tones: effects of carrier frequency discrimination and responses of primary auditory cortical neurons in the owl monkey," *Soc. Neurosci. Abstracts*, vol. 22, p. 1623, 1996.
- [8] R. E. Beitel, C. E. Schreiner, X. Wang, S. Cheung, W. M. Jenkins, and M. M. Merzenich, "Effects of psychophysical training on the entrainment of primary auditory cortical neurons to amplitude modulated tones," *Soc. Neurosci. Abstracts*, vol. 21, p. 1180, 1995.
- [9] J. L. Flanagan and R. M. Golden, "Phase vocoder," *Bell Syst. Tech.*, vol. 45, pp. 1493–1509, 1966.
- [10] M. R. Portnoff, "Implementation of the digital phase vocoder using the Fast Fourier Transform," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-24, pp. 243–248, 1976.
- [11] ———, "Time scale modification of speech based on short-time Fourier Analysis," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-29, pp. 374–390, 1981.
- [12] ———, "Short-time Fourier analysis of sampled speech," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-29, pp. 364–373, 1981.
- [13] J. B. Allen and L. R. Rabiner, "A unified approach to short-time Fourier analysis and synthesis," *Proc. IEEE*, vol. 65, pp. 1558–1564, 1977.
- [14] R. E. Crochiere, "A weighted overlap-add method for short-time Fourier analysis/synthesis," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-28, pp. 99–102, 1979.
- [15] A. V. Oppenheim and R. W. Schaffer, *Discrete-Time Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1989.
- [16] J. L. Flanagan, "Parametric coding of speech spectra," *J. Acoust. Soc. Amer.*, vol. 68, pp. 412–419, 1980.
- [17] J. B. Allen, "Short term spectral analysis, synthesis and modification by discrete Fourier transform," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-25, pp. 235–238, 1977.
- [18] J. W. Gordon and J. Strawn, "An introduction to the phase vocoder," in *Digital Audio Signal Processing: An anthology*, J. Strawn, Ed. Los Altos, CA: William Kaufmann, 1985.
- [19] T. Langhans and H. W. Strube, "Speech enhancement by nonlinear multiband envelope filtering," in *Proc. Int. Conf. Acoust., Speech, Signal Processing*, 1982, pp. 156–160.
- [20] E. Zwicker and E. Terhardt, "Analytical expressions for critical-band rate and critical bandwidth as a function of frequency," *J. Acoust. Soc. Amer.*, vol. 68, pp. 1523–1525, 1980.
- [21] K. Tremblay, N. Kraus, T. D. Carrell, and T. McGee, "Central auditory system plasticity: Generalization to novel stimuli following listening training," *J. Acoust. Soc. Amer.*, vol. 102, pp. 3762–3773, 1997.
- [22] N. Kraus, T. McGee, T. Carrell, C. King, K. Tremblay, and T. Nicol, "Central auditory system plasticity with speech discrimination," *J. Cognitive Neurosci.*, vol. 7, pp. 25–32, 1995.
- [23] D. G. Jamieson and D. E. Morosan, "Training nonnative speech contrasts in adults: Acquisition of the english /o/ -/O/ contrast by francophones," *Perception Psychophys.*, vol. 40, pp. 205–215, 1986.

- [24] ———, "Training new, nonnative speech contrasts: A comparison of the prototype and perceptual fading techniques," *Canadian J. Psychol.*, vol. 43, pp. 88–96, 1989.
- [25] S. S. Nagarajan, D. T. Blake, B. A. Wright, N. Byl, and M. M. Merzenich, "Practice-related improvements in somatosensory interval discrimination is temporally specific but generalizes across skin location, hemisphere and modality," *J. Neurosci.*, vol. 18, no. 4, pp. 1559–1570, 1998.
- [26] B. A. Wright, D. Buanomano, H. Mahncke, and M. Merzenich, "Learning and generalization of auditory temporal-interval discrimination," *J. Neurosci.*, vol. 17, pp. 3956–3963, 1997.
- [27] J. I. Alcantara, G. J. Dooley, P. J. Blamey, and P. M. Seligman, "Preliminary evaluation of a formant enhancement algorithm on the perception of speech in noise for normally hearing listeners," *Audiol.*, vol. 33, pp. 15–27, 1994.
- [28] H. T. Bunnell, "On enhancement of spectral contrast in speech for hearing impaired listeners," *J. Acoust. Soc. Amer.*, vol. 88, pp. 2546–2556, 1990.
- [29] P. M. Clarkson and S. F. Bahgat, "A real-time speech enhancement system using envelope enhancement techniques," *IEEE Trans. Electron Device Lett.*, vol. 25, pp. 1186–1189, 1989.
- [30] ———, "Envelope expansion methods for speech enhancement," *J. Acoust. Soc. Amer.*, vol. 89, pp. 1378–1382, 1991.
- [31] J. M. Kates, "Speech enhancement based on a sinusoidal model," *J. Speech Hearing Res.*, vol. 37, pp. 449–464, 1994.
- [32] V. Hohmann and B. Kollmeier, "The effect of multichannel dynamic compression on speech intelligibility," *J. Acoust. Soc. Amer.*, vol. 97, pp. 1191–1195, 1995.
- [33] B. Kollmeier and R. Koch, "Speech enhancement based on physiological and psychoacoustical models of modulation perception and binaural interaction," *J. Acoust. Soc. Amer.*, vol. 95, pp. 1593–1602, 1994.
- [34] E. Dupoux and K. Green, "Perceptual adjustment to highly compressed speech: Effects of talker and rate changes," *J. Experimental Psychol.*, vol. 23, pp. 914–927, 1997.
- [35] T. F. Quatieri and R. J. McAulay, "Shape invariant time-scale and pitch-modification of speech," *IEEE Trans. Signal Processing*, vol. 40, pp. 497–510, 1992.
- [36] R. Drullman, J. M. Festen, and R. Plomp, "Effect of reducing slow temporal modulations on speech reception," *J. Acoust. Soc. Amer.*, vol. 95, pp. 2670–2680, 1994.
- [37] ———, "Effect of temporal envelope smearing on speech reception," *J. Acoust. Soc. Amer.*, vol. 95, pp. 1053–1064, 1994.
- [38] R. Drullman, "Temporal envelope and fine structure cures for speech intelligibility," *J. Acoust. Soc. Amer.*, vol. 97, pp. 585–592, 1995.
- [39] ———, "Speech intelligibility in noise: Relative contribution of speech elements above and below the noise level," *J. Acoust. Soc. Amer.*, vol. 98, pp. 1796–1798, 1995.
- [40] R. M. Warren, K. R. Riener, J. A. Bashford, and B. S. Brubaker, "Spectral redundancy: Intelligibility of sentences heard through narrow spectral slits," *Perception Psychophys.*, vol. 57, pp. 175–182, 1995.
- [41] X. Wang, M. M. Merzenich, R. Beitel, and C. E. Schreiner, "Representation of a species-specific vocalization in the primary auditory cortex of the common marmoset: Temporal and spectral characteristics," *J. Neurophys.*, vol. 74, pp. 2685–2706, 1995.
- [42] W. F. McGee and P. Merkle, "A real-time logarithmic-frequency phase vocoder," *Comput. Music J.*, vol. 15, pp. 20–27, 1991.



**Srikantan S. Nagarajan** was born in Madras, India, in 1969. He received the Bachelor's degree in electrical engineering from the University of Madras, India. He received the Master's and Ph.D. degrees in biomedical engineering from the Applied Neural Control Laboratory, Case Western Reserve University, Cleveland, OH, in 1991 and 1995, respectively.

Since then, he has been a Postdoctoral Fellow at the Keck Center for Integrative Neuroscience at the University of California, San Francisco, where he is working on several projects related to learning and representation of spatio-temporal information processing. He is also working part-time as a Scientific Research Consultant at Scientific Learning Corporation, Berkeley, CA. His research interests include applied neural control, rehabilitation engineering, bioelectromagnetism, cortical processing and plasticity of spatio-temporal information in the nervous system, computational neuroscience, and biomedical signal processing.



**Xiaoqin Wang (M'97)** received the B.S. degree in electrical engineering from Sichuan University, Chengdu, China, in 1984, the M.S.E. degree in electrical engineering and computer science from University of Michigan, Ann Arbor, in 1986, and the Ph.D. degree in biomedical engineering from The Johns Hopkins University, Baltimore, MD, in 1991.

From 1992 to 1995, he received his postdoctoral training in neuroscience at University of California, San Francisco. He joined the Faculty of Biomedical Engineering Department at The Johns Hopkins University School of Medicine, Baltimore, MD, in 1995 and is currently an Assistant Professor of biomedical engineering and neuroscience. His research interests include information processing in the nervous system, neural mechanisms underlying speech perception and production, neurophysiology of the auditory cortex, and computational neuroscience.



**Michael M. Merzenich** was born in Lebanon, OR. He received degrees with highest honors from the University of Portland, Portland, OR. He entered a predoctoral program in Physiology at The Johns Hopkins University School of Medicine, Baltimore, MD, where he received the Ph.D. degree under the mentorship of Dr. V. Mountcastle in 1968.

After a three-year postdoctoral research period at the University of Wisconsin at Madison, he accepted a position as an Assistant Professor at the University of California, San Francisco, where he is presently a tenured rank Professor (Step V) and the Director of Research and Vice-Chairman of the Department of Otolaryngology. He is a founding member of the Keck Center for Integrative Neuroscience, San Francisco, CA, and a member of the Neuroscience, Physiology and Bioengineering training programs. His research interests include cortical network physiology, cortical contributions to learning and memory in the auditory and somatosensory systems, cortical contributions to the origins of, and bases of, recovery from stroke, and the origins and the remediation of learning disabilities, and the development of electrical stimulation prosthetic devices designed for hearing restoration for the deaf. He is the co-inventor of the commercially applied cochlear implant, and of new training instruments and methods applied for remediation of speech/language-based learning disabilities. He is also the Chief Scientific Officer at Scientific Learning Corporation, Berkeley, CA, dedicated to the development of training aids for language learning-impaired, cognitive-impaired, and motorically impaired children.



**Christoph E. Schreiner** was born in Germany in 1950. He received the Ph.D. degree in physics and the M.D. degree from the Georg-August University, Goettingen, Germany, in 1977 and 1980, respectively.

He is currently Professor of Otolaryngology at the University of California, San Francisco, and on the Faculty at the Neuroscience and Bioengineering Graduate Programs. He is the member of the Keck Center for Integrative Neuroscience, San Francisco, and the Sloan Center for Theoretical Neurobiology at the University of California, San Francisco. His current research focuses on the neuronal processing of complex sounds in the auditory cortex.



**Paul Johnston** received the Ph.D. degree in neuroscience from the University of California, San Diego in 1994.

Since then, he has been a Postdoctoral Fellow at the Keck Center for Integrative Neuroscience in the University of California, San Francisco. His research focuses on understanding basic brain mechanisms for auditory processing that underlie spoken language.



**William M. Jenkins** received the Bachelor's, Master's, and Ph.D. degrees from Florida State University, Tallahassee, and received additional training at the University of California, San Francisco.

He is currently the Vice President, Product Development at Scientific Learning Corporation, Berkeley, CA. Before joining Scientific Learning Corporation, he worked for several years as a Faculty Member in the Keck Center for Integrative Neuroscience, San Francisco, CA. His research expertise and interests include learning-based brain plasticity,

behavioral algorithms, and psychophysical methods, as well as multimedia and Internet technologies.



**Steven Miller** received the Ph.D. degree in psychology from the University of North Carolina, Greensboro, in 1991. He then had a clinical internship in neuropsychology at the Bowman Gray School of Medicine.

He is currently the Vice President of Outcomes Research at Scientific Learning Corporation, a company based in Berkeley, CA. Before joining Scientific Learning Corporation, he was a Research Associate Scientist at the Center for Molecular and Behavioral Neuroscience, Rutgers University,

Newark, NJ, where he now holds an Adjunct Faculty position. His research expertise and interests include the development of novel training programs for teaching children with language and learning problems using computers and the Internet.



**Paula Tallal** received the Ph.D. degree from Cambridge University, Cambridge, U.K., in 1973.

From 1973 to 1979, she did postdoctoral work and then was appointed as Assistant Professor at The Johns Hopkins School of Medicine, Baltimore, MD. In 1979, she joined the Faculty of the Department of Psychiatry at the University of California, San Diego, as an Assistant Professor, and was promoted to Professor in 1986. During this time, she was also Director of Research at the Child Guidance Clinic at Children's Hospital and Health Center, San Diego,

CA, and a member of the Executive Board of the University of California, San Diego/San Diego State University Joint Ph.D. Program in Clinical Psychology. She moved to Rutgers, The State University of New Jersey, in 1988, and became Professor (II) and Co-Director of the Center for Molecular and Behavioral Neuroscience. She is active on many scientific advisory boards and governmental committees, and recently served on the Diagnostic and Statistical Manual (DSM IV) Task Force work groups for both developmental language disorders and developmental learning disabilities.

Dr. Tallal has also been a member of the Institute of Medicine Task Force on Causes of Mental Disorders of Childhood and Adolescence. She recently CoFounded Scientific Learning Corporation, Berkeley, CA, where she is Executive Vice President and Chairperson of the Board of Directors.