SPEECH COMMUNICATION

ELSEVIER

Aggregate neural responses to speech sounds in the central auditory system

Nina Kraus [a,b,*], Trent Nicol [a]

[a] *Department of Communication Sciences, Northwestern University, Frances Searle Building, 2299 North Campus Drive, Evanston, IL 60208, USA*
[b] *Departments of Neurobiology and Physiology, and Otolaryngology, Northwestern University, Frances Searle Building, 2299 North Campus Drive, Evanston, IL 60208, USA*

Abstract

The fundamental complexity of speech—in both the spectral and temporal domains—elicits extensive dynamic activity from a broad neural population. Evoked potentials rely on a summation of synchronous aggregate neural activity, making them especially suitable for speech-sound investigation.

This paper summarizes research from our lab that demonstrates the efficacy of speech-evoked responses in addressing three fundamental issues. First, the neural bases of left-brain specialization to speech are investigated in an animal model. Second, studies are aimed at inferring the underlying causes of certain language-based learning disabilities. Finally, in a series of before-and-after designs, the underlying neural plasticity that accompanies directed speech-sound training is explored.
© 2002 Elsevier Science B.V. All rights reserved.

Keywords: Auditory; Neural; Perception; Speech; Perceptual learning; Evoked responses; Plasticity

1. Introduction

The following quote is taken from the mission statement of the Nature of Speech Perception workshop:

> Speech perception involves processing of many different kinds: peripheral auditory analysis, automatic feature extraction in the units of the auditory brainstem, leading to classification of words and phonemes. A major question is to what extent this processing is automatic (in which case it would presumably occur mainly in the brainstem) and to what extent it is cognitive (governed by linguistic and other high-level processing in the cortex).

This paper reviews research utilizing aggregate neural responses in order to investigate speech–sound encoding in the normal and impaired auditory system. The studies presented here have led us to conclude that the processing required for speech perception has a substantial automatic basis independent of higher-level cognitive factors.

* Corresponding author. Address: Department of Communication Sciences, Northwestern University, Frances Searle Building, 2299 North Campus Drive, Evanston, IL 60208, USA.

E-mail address: nkraus@northwestern.edu (N. Kraus).

Speech is a complex acoustic signal that places many demands on the auditory system. The healthy auditory system must be sensitive to signals with rapidly changing spectra, poor signal-to-noise ratios and fast rates of stimulation. Accurate speech representation requires underlying biological patterns of activity across broad neural populations from the periphery to the cortex. Because evoked responses depend on just that sort of synchronous activation, they are uniquely suited for examining the underlying neural bases of speech perception. Additionally, because evoked responses rely on synchronous neural firing, they provide a means for assessing a breakdown in biological timing that may be at the root of faulty speech encoding in some people with perceptual deficits.

Because of their behavior-independent nature, evoked potentials are ideal for examining the effects of subtle manipulations of the speech signal without relying on subjective behavioral responses. Furthermore, the careful characterization of evoked responses in the normal system, both in humans and in animals, can inform us about how the auditory system fails in individuals with perceptual disorders.

1.1. The approach

We take a behavioral-neurophysiological, acoustic–phonetic approach to the investigation of biological processes involved in speech–sound perception.

In humans, we assess behavioral responses to acoustically defined speech stimuli. These same stimuli are used to evoke neural responses originating from various levels along the auditory pathway (brainstem to cortex). The relationships among listeners' behavioral perception of acoustic signals, the neurophysiological representation of those same signals, and other broader measures of speech perception and language processing are examined in order to link perception of speech to underlying central physiological processes.

In animals, parallel studies utilizing the same stimuli permit direct intracranial recordings (midbrain, thalamus, cortex) to further identify neural mechanisms involved in representation of speech sounds in the brain.

By combining an examination of the perceptual and physiological responses to speech sounds, we have gained information about mechanisms underlying normal speech perception and perceptual deficits in clinical populations, particularly children with learning problems (LP). We have also begun to unravel some of the mechanisms associated with speech–sound perceptual learning.

1.2. The stimuli

Speech sounds can be characterized (at least in part) along finite and acoustically definable dimensions. The use of synthesized speech allows for precise manipulation along these dimensions and, importantly, permits discrete modifications that are difficult to control in natural speech.

1.3. The responses

Evoked responses can represent neural activity generated at different levels along the auditory pathway. These responses are elicited passively; subjects are not attending to the stimuli. Examples of evoked responses that are amenable to speech stimulation are:

Auditory brainstem response (ABR): A summation of spike discharges reflecting precise encoding of stimulus onset in the brainstem.

Frequency following response (FFR): Time-locking to periodic aspects of the speech stimulus such as fundamental frequency.

P1/N1/P2/N2: Mid- to late-latency (50–200 ms) post-synaptic cortical responses to stimulus onset.

Mismatch response (MMN): Late-latency cortical response. Requires (at least) two different stimuli. The response occurs when a train of identical stimuli is interrupted by the different stimulus.

In combination, these responses are complementary in that each differs in anatomical source, maturational time course and functional significance (Elberling et al., 1982; Sams et al., 1985; Näätänen and Picton, 1987; Sams et al., 1991; Conley et al., 1999; Møller, 1999; Ponton et al., 2000; Bellis et al., 2000).

1.4. The questions

This paper demonstrates how using speech-evoked potentials can begin to address:

- How speech sounds are represented in the brain.
- How that representation is related to perception in a normal system and in clinical populations.
- Which specific acoustic–phonetic elements are most vulnerable to mis-perception due to faulty neural encoding.
- How experience and speech–sound training alters the biology of sound perception.
- Ways in which speech–sound perception may be improved.

Specifically reviewed are studies on (1) neural lateralization for speech sounds, (2) the biological bases of certain learning disabilities, and (3) the physiological underpinnings of learning resulting from training.

2. Neural lateralization for speech sounds (King et al., 1999)

Left hemisphere lateralization for speech perception is well documented (Kimura, 1961; Geschwind, 1972; Gazzaniga, 1983; Milner, 1971). Hemispheric lateralization can arise because of the acoustic structure of a sound, and not necessarily because a sound is speech or non-speech. Converging evidence from human behavior (Tallal and Newcombe, 1978) and physiology, (Zatorre et al., 1992; Sharma et al., 1994) as well as animal behavior (Heffner and Heffner, 1986; Fitch et al., 1993) and physiology (King et al., 1999) indicates that brief, or rapidly changing acoustic cues, such as human speech and animal communication calls, elicit different responses from the left and right hemispheres.

2.1. Methods

Evoked potentials to both speech and non-speech sounds were recorded from the thalamus and cortex of 12 anesthetized guinea pigs. Tungsten 1 MΩ electrodes were placed in left and right primary pathway subdivisions of the medial geniculate (MG) nuclei. Stainless steel 10 kΩ electrodes were placed on the brain surface over left and right primary auditory cortices (AC). Simultaneous recordings were made at all four sites to a 100-ms synthesized /da/ and a 100-ms, 2-kHz tone. Stimuli were presented monaurally to both ears. Contralateral onset response amplitudes were compared.

2.2. Results

At both recording locations the tonal stimulation resulted in symmetric responses ($p > 0.05$), while the /da/ stimulus resulted in markedly larger onset amplitudes on one side. Current studies are determining which aspects of the /da/ are necessary and sufficient for eliciting an asymmetric evoked response.

2.3. Summary of results

- Lateralization of speech sounds is seen at sub-cortical levels as well as in the cortex.
- Simple, pure tone stimuli do not exhibit lateralized responses.
- Lateralization processes occur even under anesthesia; conscious attention to stimuli is likely not required.
- This is likely an evolutionary precursor to hemispheric lateralization in humans for certain complex stimuli, including speech.

3. Biological bases of learning disorders

The "Listening, Learning and the Brain" project—background: The "Listening, Learning and the Brain" project is a large-scale investigation aiming to better understand the biological bases for the sound perception deficits found in some children with LP. To achieve this, children, with and without diagnosed learning disabilities, are given a battery of psychophysical tasks (listening), standardized measures of learning and academic achievement (learning) and measures of neurophysiology (the brain).

Typically, behavioral differences between normal children and those with LP manifest themselves

when their systems are challenged by real-world listening situations such as increased speaker rate, subtle sound differences or background noise. Manipulation of these speech characteristics in an experimental setting provides a means to evaluate behavioral differences between groups and to determine to what extent these differences can be seen in physiological recordings.

Presented here are studies using speech—both behaviorally and physiologically—that challenge listeners in two ways. First, subtle variations in spectro-temporal content challenge the listener to discriminate between stop consonants, and second, a background noise masker is introduced to further challenge the auditory system. Then, cue-enhancement strategies to improve performance and neural encoding are evaluated. Finally, the same stimuli are used in an animal model to further delineate neural origins and underlying mechanisms. Moreover, it is possible to discern the extent to which the deficits seen in humans may arise from pre-attentive auditory processing mechanisms.

3.1. Fine-grained discrimination in children (Kraus et al., 1996)

The discrimination of the fine-grained acoustic differences found in certain stop consonants has been found to be particularly vulnerable in some people with LP (Tallal and Piercy, 1974; Elliott et al., 1989; Godfrey et al., 1981). To determine the extent to which this stems from underlying auditory pathway deficits, a physiological measure of preconscious discrimination, the mismatch response (MMN) was administered to two groups of school-age children. An "LP" group consisted of 91 children diagnosed with a language-based learning disability, attention deficit hyperactivity disorder, or both. A control group consisted of 90 children in the same age range as the LP group, but with no known learning problems.

3.1.1. Behavioral methods

To determine the children's behavioral fine-grained speech discrimination ability, two consonant–vowel (CV) continua, comprising five-formant, 100-ms synthetic speech syllables, were created. The first continuum employed a place of articulation contrast that has been shown to be more difficult to perceive than most other phonetic contrasts (e.g., Miller and Nicely, 1955): /da/ to /ga/ which varied only in the onset frequency of the third formant. A control continuum, /ba/ to /wa/, differed only in duration of the first and second formant transitions. A task to determine just noticeable differences (JND) on both contrasts was administered to all children (Taylor and Creelman, 1967; Carrell et al., 1999).

3.1.2. Physiological methods

Stimuli: Mismatch responses were recorded in 42 children. Two age-matched groups consisted of children with equally good /ba–wa/ discrimination and either good or poor /da–ga/ discrimination. The /ba–wa/ pair chosen (5 ms formant transition duration difference) was one that was discriminable by children in both groups. The /da–ga/ pair chosen (80 Hz difference in onset frequency of third formant) was discriminable by children in the control, but not the experimental group.

General physiological protocol: Evoked response recording for all human studies presented in this paper shared the following standard protocol. Subjects were seated comfortably in an acoustically shielded booth. Speech stimuli were delivered monaurally to the right ear through insert earphones. The left ear was left unoccluded in order to allow subjects to hear a movie soundtrack that was kept at a low volume (∼40 dB SPL). The movie, chosen by the subject, was used to minimize any effects of drowsiness. Electrodes were silver/silver-chloride and had impedances of less than 5 kΩ. The nose was common reference and the forehead served as ground. Active electrodes were Fz, Cz, Pz, F3, F4, A1, A2 and a pair just anterior to T3 and T4 that we term TL and TR. Two electrodes were placed near the left eye to enable rejection of trials contaminated with eyeblinks.

Response recording: Stimuli were presented at an intensity of 75 dB SPL. In addition to the frequent and rare, or oddball, stimulation protocol, a separate 'rare-alone' recording was undertaken so that responses to the same stimulus could be compared whether it appeared as an oddball or by itself. This served to minimize any inherent response differences that may exist between the

stimuli. All measures (latency, amplitude, duration, area) of MMN were calculated on a difference wave computed by subtracting the rare-alone response from the rare (oddball) response.

3.1.3. Behavioral results

LP children performed significantly more poorly than the controls on the /da–ga/ continuum ($p < 0.001$). Children in the two groups performed equivalently well on the /ba–wa/ continuum, indicating that the deficit seen in the LP children to /da–ga/ was purely a perceptual deficit and not an inability to perform the task.

3.1.4. Physiological results

Both groups showed robust mismatch responses to the /ba–wa/ pair, but, mirroring their perceptual difficulties, the experimental group had very small or absent MMNs to the /da–ga/ stimulus pair, while the control group had strong MMNs to /da–ga/ ($p < 0.003$). This was the first demonstration that perceptual speech processing deficits in LP children have a basis in faulty, pre-attentive neural encoding of signals along the auditory pathway, specifically in auditory cortex.

3.1.5. Summary of results

- Children with diagnosed LP have difficulty perceiving *certain* speech contrasts that are key to the proper perception of speech sounds. Stimuli with rapid spectral changes are particularly vulnerable to misperception.
- LP children exhibited degraded MMN responses to /da–ga/ and robust responses to /ba–wa/, indicating that faulty neural encoding of certain speech signals is implicated in poor behavioral speech perception. This deficit is not contingent upon attention because it is apparent in passively elicited neural responses.

3.2. Brainstem timing deficits to speech in children (King et al., 2002)

The cortical deficits to speech sounds seen in LP children, described in Section 3.1, led us to examine whether this population differed from normal children in even more basic neural responses

recorded from lower in the auditory pathway. The auditory brainstem response (ABR) to both a click and the speech sound /da/ was recorded in 33 normal and 54 LP children. Cortical P1/N1/P2/N2 responses to /da/, presented in both quiet and background noise, also were recorded.

3.2.1. Methods

Brainstem response recording: The endpoint /da/ from the continuum described in Section 3.1.1 was stripped of its final 60 ms vowel and presented monaurally to the right ear at a rate of 11.1 s^{-1}. Sixty ms of post-stimulus activity was recorded to permit examination of both the onset response (wave V) and the subsequent FFR evoked by the harmonic structure of /da/. As a control, a 100 μs click was presented at a rate of 31.1 s^{-1}. Both stimuli were presented at alternating polarity to negate the confound of the cochlear microphonic. The recording site was Cz, referenced to right earlobe.

Cortical response recording: The same /da/ was presented monaurally to the right ear both in quiet and in continuous white background noise (SNR = 0 dB) at a rate of 1.7 s^{-1}. The recording site was Cz, referenced to nose.

3.2.2. Analysis

Brainstem responses: To the click stimulus, latency of wave V was documented. To the /da/, three peaks were found to be consistent. An onset response, at a latency of about 7.5 ms, was followed by two later FFR peaks at 17.7 and 39.5 ms. For all response peaks, a normal limit was defined as the mean plus one standard deviation of the normal subjects.

Cortical responses: For responses recorded in both quiet and background noise, latencies and amplitudes for P1/N1/P2/N2 were marked. A cross correlation was performed between the response in quiet and the response in background noise. The maximum correlation and the time shift at which it occurred were recorded.

3.2.3. Results

Brainstem responses: There were no differences between the normal and LP groups in wave V latency to the click stimulus. The onset response to

the /da/ was significantly later for LP than normal subjects ($p < 0.03$). Of the 54 LP subjects, 20 fell outside the normal range in onset response latency. The same 20 subjects also had delays in the two FFR peaks ($p < 0.01$ for both). To the click, these subjects did not fall outside the normal range. Of the 34 LP subjects that were within the normal range for onset response latency, only two had delayed FFR peak latencies.

Relationship between brainstem and cortical responses: The twenty LP subjects with delays in their ABR latencies had significantly lower cortical response quiet-to-noise correlations ($p < 0.02$) than the 34 LPs with normal ABR latencies. That is, waveform morphology was degraded to a greater extent by the addition of background noise in children with abnormal brainstem responses.

3.2.4. Summary of results
- Some LP children have delayed brainstem responses to speech stimuli.
- These delays are seen in both onset and frequency-following components of the response.
- The same children exhibiting abnormal brainstem response encoding have degraded cortical responses in a measure that is sensitive to timing disruptions in background noise.

3.3. Effects of background noise and cue enhancement in children (Cunningham et al., 2001; Wible et al., 2002)

Children with speech–sound perception deficits have difficulty perceiving speech in noisy listening situations. We hypothesized that in these children, the basic neural representation of sound would be particularly disrupted in noise. Thus brainstem and cortical evoked potentials were recorded in quiet and continuous background noise in both LP children and normal controls. Additionally, cue-enhancement strategies that are known to be effective (Picheny et al., 1985) were examined physiologically to gauge whether the improvement seen behaviorally is accompanied by a change in the evoked response to the same stimuli.

Subjects were normal children and age-matched LP children who performed significantly worse than normal children on measures of auditory processing (Woodcock and Johnson, 1977, 1989), reading, spelling (Wilkinson, 1993), and fine-grained discrimination along the /da–ga/ continuum described in Section 3.1.1.

3.3.1. Behavioral methods
Four vowel–consonant–vowel (VCV) continua were created. The baseline (conversational) continuum /ada/ to /aga/ was 40 steps long and varied only in the $F3$ onset frequency of the consonant. The three other continua employed the following cue-enhancement strategies: (a) consonant release burst intensity was amplified by 10 dB, (b) the stop gap was lengthened by 80 ms, and (c) a combination of both strategies. All continua differed only in the consonant's $F3$ onset frequency and were delivered binaurally in continuous background white noise (SNR $= +5$ dB).

3.3.2. Behavioral results
In background noise, the LP subjects had significantly worse fine-grained discrimination than the normal children on the baseline continuum ($p < 0.001$). With the cue-enhanced continua, perception in the LP children was improved to the level of the normal subjects. Of the cue-enhancement strategies, increasing the release burst intensity was found to have more effect than lengthening the stop gap.

3.3.3. Neurophysiological methods
Because the lengthened stop gap provided little improvement in behavioral discrimination, it was decided that the evoked response recording would concentrate on the release burst enhancement alone. This allowed the initial /a/ to be stripped from the stimuli to enable faster data collection.

ABR, FFR and P1/N1/P2/N2 cortical potentials were obtained in response to the endpoint /da/ stimulus from the baseline continuum and from the enhanced release burst continuum. Stimuli were presented at a rate of 11 s^{-1} monaurally to the right ear at 80 dB SPL both in quiet and with a continuous background noise masker (SNR $= +5$ dB). The active electrode was Cz. See Section 3.1.2

General physiological protocol for additional details.

Onset latency of ABR wave V, frequency content and stimulus-to-response correlation of the FFR were determined. Latencies, amplitudes and correlations of the cortical P1/N1/P2/N2 responses were compared between the subject groups for each stimulus condition.

3.3.4. Neurophysiological results

ABR, FFR and P1/N1/P2/N2 responses were equivalent in the two groups when the /da/ was presented in quiet. The addition of background noise resulted in LP subjects showing (1) a reduction of energy in the 250–750 Hz range of the FFR ($p < 0.01$); (2) poorer FFR timing as measured by stimulus-to-response correlation; and (3) reduced P2-to-N2 amplitude ($p < 0.01$). Normal control subjects' responses were stable on these measures with the addition of background noise.

In a separate study, responses to speech stimuli that were rapidly presented in blocks of four, both in quiet and background noise, were compared between normal ($n = 12$) and LP ($n = 13$) children. The cortical response to the first stimulus was correlated with the response to the fourth stimulus over a latency range that encompassed the P2/N2 time window. This provided an index of how response timing was affected by stresses of stimulus repetition and background noise. In quiet, both the normal and LP groups' responses had equivalently high correlations. In background noise, however, there was a wide range of response correlations in the LP group, while the normal group's correlations were all above 0.75. This wide range of response correlations in the LPs was significantly related to their /da–ga/ JND scores as well as standardized measures of auditory processing and spelling.

Cue-enhanced stimuli: When /da/ is presented as a cue-enhanced signal, LP children exhibited normal cortical responses in noise. This implies that the LP group's discrimination improvement to normal levels with cue-enhanced stimuli may have a direct link to more accurate neural representation of acoustic events.

3.3.5. Summary of results

- In background noise, behavioral discrimination of LP children suffers to a greater extent than in normal children.
- Background noise disrupts physiological responses of brainstem and cortical origin to speech stimuli more in LP than normal children.
- Cue enhancement strategies are effective in helping LPs achieve normal speech–sound perception.
- Cortical responses in LP children become "normal" with cue-enhanced speech stimuli.

3.4. Effects of background noise and cue enhancement in an animal model (Cunningham et al., 2002)

In order to investigate the underlying sources of the physiological deficits seen in LP children in experiment 3.3, parallel studies using guinea pigs were undertaken. First, the effects of background noise were studied, and then the benefits of cue-enhancement strategies were assessed. Responses from three recording sites along the auditory pathway were evaluated to investigate contributions of midbrain, thalamus and cortex.

3.4.1. Methods

Stimuli: The four endpoint stimuli of the continua from experiment 3.3 were used. Because of lesser time constraints in animal testing, it was possible to use the entire VCV so that the efficacy of both cue-enhancement strategies (lengthened stop gap, increased release burst intensity) could be evaluated. Stimuli were delivered monaurally to the right ear, both in quiet and with continuous background noise (SNR = 0 dB).

Response recording: Responses were recorded from midbrain, the central division of the inferior colliculus (IC); thalamus, the ventral division of the MG nucleus; and over the primary AC.

Analysis: Analyses focused on the response to the /da/ portion of the /ada/ stimulus. Onset and offset response amplitudes, spectral content of the sustained response (FFR), and precision of phase-locking (stimulus-to-response correlation) of the sustained response were measured from all recording locations.

3.4.2. Results

Background noise: Background noise disrupted the onset response to /da/ to a greater extent than the vowel portion of the stimulus ($p < 0.02$). This disruption was more pronounced at the cortical level than at the two subcortical locations. For the sustained response in IC, noise reduced the magnitude of low-frequency spectral components ($p < 0.02$) whereas higher-frequency components remained unchanged, mirroring findings in normal children using the same stimuli.

Cue enhancement: First, the separate contributions of lengthened stop gap duration and increased release burst intensity of the syllable /ada/ were evaluated. The amplitude of the onset response to the consonant /da/ increased as either the stop gap duration or the burst intensity alone were increased. Combined manipulations maximally increased the amplitude of the onset response ($p < 0.02$ at all anatomical areas) and were not a simple linear sum of the effect of each manipulation alone. This improvement was greatest at AC. In background noise, enhanced speech stimuli elicited an onset response to /d/ which was absent with unenhanced signals. Further analysis of the latter portions of the response indicated that (as in normal children) cue enhancements did not affect the representation of sustained or offset responses.

3.4.3. Summary of results

- Onset responses are more affected by background noise than FFRs to speech stimuli.
- Cue-enhancement strategies benefit onset responses more than FFRs.
- These onset-response effects are more pronounced at the cortex than in midbrain or thalamus.
- In response to noise and cue enhancement, the neurophysiological patterns seen in IC sustained response mirror those seen in normal children.
- The particular susceptibility of the auditory cortex to manipulations of noise and cue enhancement may be related to the finding that in LP children, the effects of noise and cue enhancement were more dramatic in cortical than subcortical responses.

4. Perceptual learning

Speech perception abilities in humans are modifiable both by long-term experience with one's native language (Mehler et al., 1978; Werker et al., 1992; Kuhl et al., 1992; Jusczyk et al., 1993; Näätänen et al., 1997; Dehaene-Lambertz and Baillet, 1998) and short-term auditory training in a laboratory or clinical environment (Pisoni et al., 1982; Shankweiler et al., 1995; Tallal et al., 1996; Merzenich et al., 1996; Bradlow and Pisoni, 1999). Considered here are physiological responses compared prior to and after auditory training.

Animal experiments have amply demonstrated sensory cortex restructuring with training (Jenkins et al., 1990; Merzenich et al., 1990; Recanzone et al., 1993). In humans, evoked potentials can be used to assess underlying neural changes associated with behavioral learning as reviewed below.

4.1. Perception training in adults (Kraus et al., 1995; Tremblay et al., 2001)

Because the mismatch response (MMN) occurs without attention and can be elicited by small stimulus differences, it can be used to study training-associated improvement in discriminationthresholds. In other words, one can measure whether, following training, the auditory system responds to a formerly undetectable difference, separate from a subject's conscious perception.

4.1.1. Methods

Thirteen adult subjects were trained to discriminate between two syllables that differed by 15 and 175 Hz in the onset frequencies of $F2$ and $F3$, respectively. Prior to training, in a /da/ vs /ga/ identification task, both syllables were identified as /da/ 100% and 96% of the time. A training regimen consisted of a two-alternative same-different task with feedback. There were six training sessions over a one-week period. All subjects performed at chance at the beginning of the training. MMNs were recorded (see Section 3.1.2 *General physiological protocol* for additional details) to the same stimulus pair prior to and upon completion of the training regimen.

4.1.2. Results

By the final session, six of the thirteen subjects were able to discriminate above chance. A one-month follow-up revealed that the improvement persisted. Mismatch responses, prior to behavioral training, were present in ten subjects. After training, MMNs were present in all thirteen subjects. Moreover, there were significant increases in response magnitude ($p < 0.01$ for duration, amplitude and area). In an associated study (Tremblay et al., 2001), changes in N1/P2 amplitude were also observed following training ($p < 0.001$).

The results indicate that while there is not a one-to-one correlation between these physiological measures and performance, as a group, enhanced cortical responses to the stimuli used in training were mirrored by improvement in perception. Moreover, active training to discriminate a particular stimulus pair can be reflected passively in the auditory system's response to the same stimulus pair in a non-attentive setting.

4.1.3. Summary of results
- Discrimination training of fine-grained speech signals results in modified cortical neural representation of stimulus differences.

4.2. Time course of learning in adults (Tremblay et al., 1998)

In order to further scrutinize how behavioral perceptual improvements are reflected physiologically, an investigation was undertaken to examine whether physiological changes preceded the behavioral speech perception improvement.

4.2.1. Methods

Monolingual English speakers were trained to distinguish between two pre-voiced consonant–vowel syllables. This distinction is not a relevant cue in the English language. The subjects, prior to training, heard both syllables as /ba/, that is, they failed to detect the pre-voicing. Four sessions over a period of ten days were employed to train subjects to label a syllable with a pre-voicing of 10 ms as /ba/ and a syllable with pre-voicing of 20 ms as /mba/.

Mismatch response sessions (see Section 3.1.2 *General physiological protocol* for additional details) using this same stimulus pair were recorded prior to training, on days following each training session, and upon completion of the entire training regimen. Onset latencies and amplitudes, durations, and areas of the response were noted for each session.

4.2.2. Results

By the end of the training regimen, nine of ten subjects learned to discriminate between the two pre-voiced syllables. All ten subjects exhibited changes in MMN. The general response pattern was an increase in both duration ($p < 0.001$) and area ($p < 0.001$) of the response with an accompanying decrease in onset latency ($p < 0.05$).

For all subjects, the change in the MMN occurred *immediately*, on the day following the first training session. However, four of the subjects required additional training to behaviorally distinguish between the two syllables. That is, they exhibited changes in the evoked responses prior to successful behavioral discrimination. Behavioral and evoked response changes were simultaneous for the remaining subjects.

4.2.3. Summary of results
- For this stimulus distinction, neurophysiological changes occurred for all subjects shortly after the onset of the training regimen—indicating that the auditory system had begun to encode the acoustic difference between the two stimuli.
- The time course of behavioral learning was more variable—for some subjects concurrent with physiological change, for some *after* physiological change, for some not at all.

It appears that a population of cells contributing to the mismatch response began to differentiate their firing patterns for the two contrasting stimuli almost immediately. This "fast learning" (Polat and Sagi, 1988) process must have occurred preattentively because in the initial stages of training, not all subjects were aware of the salient feature that separated the stimuli. Following the

automatic pre-attentive learning, a later, slow component took place.

4.3. Generalization of training in adults (Tremblay et al., 1997)

Following successful training on a particular speech contrast, learning may generalize to novel stimuli with similar acoustic properties (McClaskey et al., 1983). From a clinical standpoint, generalization is important because it is the basis of many targeted auditory training regimens in use today. Evoked responses were used to investigate neural plasticity that underlies generalization of learning.

4.3.1. Methods
In a design similar to that of Section 4.2, native English-speaking subjects were trained to discriminate between an unfamiliar stimulus pair, /mba/ and /ba/, differing in duration of prevoicing. Prior to and after training, subjects were evaluated on two behavioral tasks, without feedback to establish their baseline performance. The tasks were a same-different discrimination task and a labeling task. Mismatch responses were recorded to the same stimulus pair (see Section 3.1.2 *General physiological protocol* for additional details). Additionally, the same behavioral and physiological tests were administered using a novel speech pair, /nda/ and /da/. This stimulus pair shares the same prevoicing duration distinction, but uses alveolar rather than bilabial articulation.

The training regimen consisted of nine sessions in five days of behavioral training *only* on the /mba–ba/ pair. The training sessions used a labelling task with feedback.

4.3.2. Results
Comparisons of the pre- and post-training behavioral scores, revealed improvements to both the trained ($p < 0.05$) and untrained stimulus pairs ($p < 0.05$). As expected, MMNs following training were larger in duration ($p < 0.01$) and area ($p < 0.05$) for the trained /mba–ba/ pair. Moreover, generalization to the novel stimulus pair, /nda–da/, was reflected in increased response duration ($p < 0.01$) and area ($p < 0.05$).

4.3.3. Summary of results
- Following training, learning generalized to a novel but similar stimulus.
- Following training, a neural change was observed in response to both the trained and novel stimulus pairs.

4.4. Neural and perceptual consequences of commercial training in children (Hayes et al., 2001; King et al., 2002)

Treatment strategies for dealing with perceptual problems—particularly, computer-based auditory training programs—have become a focus in neuroscience research (Orton, 1937; Tallal et al., 1998; Morrison, 1998; Diehl, 1999). However, the efficacy of these programs is not uniform across children, and it is unclear which kind of training, for which profile of deficits, results in perceptual improvement. Studying children who undergo this training provides insights into the neurophysiological and perceptual changes associated with perceptual learning. Because of the heterogeneity of this population, it is important to determine *which* children might benefit from training, and how training may alter the neural representation of sound at various levels of the auditory pathway.

4.4.1. Methods
Twenty-five subjects with auditory-based LP were tested on a number of behavioral and physiological measures prior to and after their commercial training (Earobics) regimens. The training regimen consisted of independently administered directed exercises that incorporate phoneme discrimination, auditory memory, sequencing and attention, rhyming and sound blending skills.

The cognitive battery that was used before and after commercial training included subtests of the Woodcock–Johnson Psycho-Educational Battery and the Wide Range Achievement Test. Behavioral listening tasks were administered including sentence perception in noise and JNDs for various CV syllables in quiet and noise. The physiological regimen consisted of ABR recording to /da/ both in quiet and in noise, intrinsic cortical P1/N1/P2/N2 responses to /da/ in quiet and noise, and MMN

to /da–ga/ in quiet (see Section 3.1.2 *General physiological protocol*). Non-trained controls ($n = 16$) were also tested and retested after a similar interval.

4.4.2. Results

A range of measures of learning and speech perception showed improvement following training. The Woodcock–Johnson subtest scores that improved most included those associated with phonetic awareness and listening comprehension. Behaviorally, some of the trained subjects showed improved JNDs on an enhanced /da–ga/ continuum in noise. There were some changes in cortical evoked responses. Intrinsic P1/N1/N2/P2 responses took on a more mature-looking pattern in quiet following training, while the amount of degradation with the addition of noise diminished. Topographic patterns of MMN changed. The overall response area shifted more to the left hemisphere following training. Children with impaired brainstem encoding showed the greatest training-associated improvements: both speech–sound discrimination and resistance of cortical responses to the degrading effects of background noise were observed.

4.4.3. Summary of results

- Children who underwent focused auditory training exhibited behavioral improvements on standardized tests that tap into auditory-based skills.
- Following training, neural representation of speech-sounds altered at an automatic, pre-attentive level.

5. Conclusions

Speech-evoked neurophysiological responses, viewed in conjunction with behavioral tasks utilizing the same stimuli, provide an approach for establishing relationships between perceptual abilities and underlying central physiological processes. Additionally, using the same stimuli in an animal model provides a means of further localizing processes along the auditory pathway that contribute to the cortical representation of complex sounds such as speech.

The experimental approach described here, which includes evoked responses to speech, provides a framework for understanding some of the physiological and psychoacoustical processes contributing to speech perception. The spectral and temporal complexity of a speech signal elicits patterns of time-synchronized activity from a broad neural population. Maintaining accurate timing, as the speech stimulus is degraded by the effects of noise and other challenging listening situations, contributes to successful speech perception. This approach has elucidated why certain speech sounds are more vulnerable to perceptual and encoding disruption than others. Moreover, we are learning which aspects of auditory pathway encoding are most affected by adverse listening conditions such as noise, and what processes underlie the benefits brought about by cue-enhancement.

The results of the studies discussed here provide evidence that key processes, essential for the accurate perception of speech, occur automatically. Some of these processes such as lateralization occur subcortically in anesthetized animals.

In summary, an acoustic–phonetic experimental approach, combining physiological recording with behavioral evaluation, allows us to begin to link single-neuron physiology in animals, speech-elicited neural responses in people, and speech perception. Importantly, associations between speech–sound perception and underlying neural processes can be made in normal and clinical populations. Furthermore, these data indicate that the representation of speech sounds is plastic and can be modified by listening training. Rehabilitative strategies involving enhancing speech contrasts or focused listening training may prove efficacious in some populations that exhibit speech perception deficits of central auditory pathway origin.

References

Bellis, T.J., Nicol, T., Kraus, N., 2000. Aging affects hemispheric asymmetry in the neural representation of speech sounds. J. Neurosci. 20, 791–797.

Bradlow, A.R., Pisoni, D.B., 1999. Recognition of spoken words by native and non-native listeners: talker-, listener-, and item-related factors. J. Acoust. Soc. Amer. 106, 2074–2085.

Carrell, T.D., Bradlow, A.R., Nicol, T.G., Koch, D.B., Kraus, N., 1999. Interactive software for evaluating auditory discrimination. Ear. Hear. 20, 175–176.

Conley, E.M., Michalewski, H.J., Starr, A., 1999. The N100 auditory cortical evoked potential indexes scanning of auditory short-term memory. Clin. Neurophysiol. 110, 2086–2093.

Cunningham, J., Nicol, T., Zecker, S.G., Kraus, N., 2001. Neurobiologic responses to speech in noise in children with learning problems: deficits and strategies for improvement. Clin. Neurophysiol. 112, 758–767.

Cunningham, J., Nicol, T., King, C.D., Zecker, S.G., Kraus, N., 2002. Effects of noise and cue enhancement on neural responses to speech in auditory midbrain, thalamus and cortex. Hear. Res. 169, 97–111.

Dehaene-Lambertz, G., Baillet, S., 1998. A phonological representation in the infant brain. NeuroReport 9, 1885–1888.

Diehl, S., 1999. Listen & learn? A software review of Earobics, Lang. Speech Hear. Serv. Schools 30, 108–116.

Elberling, C., Bak, C., Kofoed, B., Lebech, J., Saermark, K., 1982. Auditory magnetic fields: source location and 'tonotopical organization' in the right hemisphere of the human brain. Scand. Audiol. 11, 61–65.

Elliott, L.L., Hammer, M.A., Scholl, M.E., 1989. Fine-grained auditory discrimination in normal children and children with language-learning problems. J. Speech Hear. Res. 32, 112–119.

Fitch, R.H., Brown, C.P., Tallal, P., 1993. Left hemisphere specialization for auditory temporal processing in rats. Ann. NY Acad. Sci. 682, 346–347.

Gazzaniga, M.S., 1983. Right hemisphere language following brain bisection. A 20-year perspective. Amer. Psychol. 38, 525–537.

Geschwind, N., 1972. Language and the mind. Sci. Amer. 226, 76–83.

Godfrey, J.J., Syrdal-Lasky, A.K., Millay, K.K., Knox, C.M., 1981. Performance of dyslexic children on speech perception tests. J. Exp. Child Psychol. 32, 401–424.

Hayes, E., Nicol, T., Zecker, S., Kraus, N., 2001. Auditory processing and neural plasticity with auditory perceptual training in children with learning problems. Assoc. Res. Otolaryngol, Abstr., 22045.

Heffner, H., Heffner, R., 1986. Effect of unilateral and bilateral auditory cortex lesions on the discrimination of vocalization by Japanese macaques. J. Neurophysiol. 56, 683–701.

Jenkins, W.M., Merzenich, M.M., Ochs, M.T., Allard, T., Guic-Robles, E., 1990. Functional reorganization of primary somatosensory cortex in adult owl monkeys after behaviorally controlled tactile stimulation. J. Neurophysiol. 63, 82–104.

Jusczyk, P.W., Cutler, A., Redanz, N.J., 1993. Infants' preference for the predominant stress patterns of English words. Child Dev. 64, 675–687.

Kimura, D., 1961. Cerebral dominance in the perception of verbal stimuli. Can. J. Psychol. 15, 166–171.

King, C., Nicol, T., McGee, T., Kraus, N., 1999. Thalamic asymmetry is related to acoustic signal complexity. Neurosci. Lett. 267, 89–92.

King, C., Cunningham, J., Nicol, T., Warrier, C., Kraus, N., 2002. Deficits in auditory brainstem encoding of speech sounds in children with learning problems. Neurosci. Lett. 319, 111–115.

Kraus, N., McGee, T., Carrell, T., King, C., Tremblay, K., 1995. Central auditory system plasticity associated with speech discrimination training. J. Cogn. Neurosci. 7, 27–34.

Kraus, N., McGee, T.J., Carrell, T.D., Zecker, S.G., Nicol, T.G., Koch, D.B., 1996. Auditory neurophysiologic responses and discrimination deficits in children with learning problems. Science 273, 971–973.

Kuhl, P.K., Williams, K.A., Lacerda, F., Stevens, K.N., Lindblom, B., 1992. Linguistic experience alters phonetic perception in infants by 6 months of age. Science 255, 606–608.

McClaskey, C.L., Pisoni, D.B., Carrell, T.D., 1983. Transfer of training of a new linguistic contrast in voicing. Percept. Psychophys. 34, 323–330.

Mehler, J., Bertoncini, J., Barriere, M., 1978. Infant recognition of mother's voice. Perception 7, 491–497.

Merzenich, M.M., Recanzone, G.H., Jenkins, W.M., Grajski, K.A., 1990. Adaptive mechanisms in cortical networks underlying cortical contributions to learning and nondeclarative memory. Cold Spring Harb. Symp. Quant. Biol. 55, 873–887.

Merzenich, M.M., Jenkins, W.M., Johnston, P., Schreiner, C., Miller, S.L., Tallal, P., 1996. Temporal processing deficits of language-learning impaired children ameliorated by training. Science 271, 77–81.

Miller, G.A., Nicely, P.E., 1955. An analysis of perceptual confusions among some English consonants. J. Acoust. Soc. Amer. 27, 338–352.

Milner, B., 1971. Interhemispheric differences in the localization of psychological processes in man. Br. Med. Bull. 27, 272–277.

Møller, A.R., 1999. Neural mechanisms of BAEP. Electroencephalogr. Clin. Neurophysiol. Suppl. 49, 27–35.

Morrison, S., 1998. Computer applications: Earobics Pro. Child Lang. Teaching Ther. 14, 279–284.

Näätänen, R., Picton, T., 1987. The N1 wave of the human electric and magnetic response to sound: a review and an analysis of the component structure. Psychophysiology 24, 375–425.

Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., Vainio, M., Alku, P., Ilmoniemi, R.J., Luuk, A., Allik, J., Sinkkonen, J., Alho, K., 1997. Language-specific phoneme representations revealed by electric and magnetic brain responses. Nature 385, 432–434.

Orton, S.T., 1937. Reading, Writing and Speech Problems in Children. New York, Norton.

Picheny, M.A., Durlach, N.I., Braida, L.D., 1985. Speaking clearly for the hard of hearing I: Intelligibility differences between clear and conversational speech. J. Speech Hear. Res. 28, 96–103.

Pisoni, D.B., Aslin, R.N., Perey, A.J., Hennessy, B.L., 1982. Some effects of laboratory training on identification and discrimination of voicing contrasts in stop consonants. J. Exp. Psychol. Human Percept. Perform 8, 297–314.

Polat, U., Sagi, D., 1988. Plasticity of spatial interactions in early vision. In: Julesz, B., Kovacs, I. (Eds.), Maturational Windows and Adult Cortical Plasticity. Addison-Wesley, New York, pp. 111–122.

Ponton, C.W., Eggermont, J.J., Kwong, B., Don, M., 2000. Maturation of human central auditory system activity: evidence from multi-channel evoked potentials. Clin. Neurophysiol. 111, 220–236.

Recanzone, G.H., Schreiner, C.E., Merzenich, M.M., 1993. Plasticity in the frequency representation of primary auditory cortex following discrimination training in adult owl monkeys. J. Neurosci. 13, 87–103.

Sams, M., Paavilainen, P., Alho, K., Näätänen, R., 1985. Auditory frequency discrimination and event-related potentials. Electroencephalogr. Clin. Neurophysiol. 62, 437–448.

Sams, M., Kaukoranta, E., Hämäläinen, M., Näätänen, R., 1991. Cortical activity elicited by changes in auditory stimuli: different sources for the magnetic N100m and mismatch responses. Psychophysiology 28, 21–29.

Shankweiler, D.P., Crain, S., Katz, L., Fowler, A.E., Liberman, A.M., Brady, S., Thorton, R., Lundquist, E., Dreyer, L., Fletcher, J.M., Stuebing, K.K., Shaywitz, S.E., Shaywitz, B.A., 1995. Cognitive profiles of reading-disabled children: Comparison of language skills in phonology morphology and syntax. Psychol. Sci. 6, 149–156.

Sharma, A., Kraus, N., Nicol, T., Carrell, T., McGee, T., 1994. Physiological bases of pitch and place of articulation perception: a case study. J. Acoust. Soc. Amer. 95, 3011.

Tallal, P., Newcombe, F., 1978. Impairment of auditory perception and language comprehension in dysphasia. Brain Lang. 5, 13–34.

Tallal, P., Piercy, M., 1974. Developmental aphasia: Rate of auditory processing and selective impairment of consonant perception. Neuropsychologia 12, 83–93.

Tallal, P., Miller, S.L., Bedi, G., Byma, G., Wang, X., Nagarajan, S.S., Schreiner, C., Jenkins, W.M., Merzenich, M.M., 1996. Language comprehension in language-learning impaired children improved with acoustically modified speech. Science 271, 81–84.

Tallal, P., Merzenich, M.M., Miller, S., Jenkins, W., 1998. Language learning impairments: integrating basic science technology and remediation. Exp. Brain Res. 123, 210–219.

Taylor, M.M., Creelman, C.D., 1967. PEST: Efficient estimates on probability functions. J. Acoust. Soc. Amer. 41, 782–787.

Tremblay, K., Kraus, N., Carrell, T.D., McGee, T., 1997. Central auditory system plasticity: generalization to novel stimuli following listening training. J. Acoust. Soc. Amer. 102, 3762–3773.

Tremblay, K., Kraus, N., McGee, T., 1998. The time course of auditory perceptual learning: neurophysiological changes during speech-sound training. NeuroReport 9, 3557–3560.

Tremblay, K., Kraus, N., McGee, T.J., Ponton, C.W., Otis, B., 2001. Central auditory plasticity: changes in the N1-P2 complex following speech-sound training. Ear. Hear. 22, 79–90.

Werker, J.F., Frost, P.E., McGurk, H., 1992. La langue et les Levres: Cross-language influences on bimodal speech perception. Can. J. Psychol. 46, 551–568.

Wible, B., Nicol, T., Kraus, N., 2002. Abnormal neural encoding of repeated speech stimuli in noise in children with learning problems. Clin. Neurophysiol. 113, 485–494.

Wilkinson, G., 1993. Wide Range Achievement Test-3. Jastak Assoc, Willington, DE.

Woodcock, R., Johnson, M., 1977. Woodcock–Johnson Psycho-Educational Battery—Revised. DLM Teaching Resources, Allen, DX.

Woodcock, R., Johnson, M., 1989. Woodcock–Johnson Psycho-Educational Battery—Revised. DLM Teaching Resources, Allem, TX.

Zatorre, R.J., Evans, A.C., Meyer, E., Gjedde, A., 1992. Lateralization of phonetic and pitch discrimination in speech processing. Science 256, 846–849.