Chapter 66

# Speech-sound encoding: physiological manifestations and behavioral ramifications

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

[a] *Department of Communication Sciences, Northwestern University, Frances Searle Building, 2240 Campus Drive, Evanston, IL 60208 (USA)*
[b] *Department of Neurobiology and Physiology; Otolaryngology, Northwestern University, Frances Searle Building, 2240 Campus Drive, Evanston, IL 60208 (USA)*

## 1. Introduction

In order to fully process and understand speech, it is necessary that its neural encoding remain intact as the signal is transduced from the eighth nerve to the auditory cortex. Imperfections in this transduction may occur in the brainstem, midbrain, thalamus or the cortex itself. Unsatisfactory speech perception is experienced by everyone to some degree when exposed to noisy, or otherwise adverse, listening conditions. However, there are those who, despite normal peripheral hearing thresholds, experience speech perception difficulties even in relatively non-challenging listening conditions. In these populations, such as the elderly and individuals with auditory-based learning disabilities, there is an interest in identifying and localizing the defect and, more importantly, taking corrective steps to improve speech perception.

The speech signal, unlike most other naturally occurring sounds, is composed of harmonically rich components that rapidly change in frequency. The timing and direction of the frequency sweeps and the relative spacing of the harmonic components combine to form the consonant and vowel sounds that are relevant to language. This complex spectro-temporal structure requires an exquisitely well-coordinated ensemble neural response for accurate encoding. This coordinated timing of neural ensembles, in turn, is a property that is suited for measurement by evoked potential averaging.

This report will review research demonstrating both cortically and subcortically how speech sounds are encoded in normal and impaired populations. Perceptual improvements that arise from both stimulus manipulation and auditory training also are evident in the aggregate neural responses, revealing that basic encoding of sound structure in the nervous system can be altered. In some cases, success of a training regimen may be predicted by how speech is encoded by the auditory pathway before training.

## 2. Measuring speech-sound encoding

Described here are three examples of physiological manifestations of speech-sound encoding that inform us about the normal encoding process and are

* *Correspondence to*: Nina Kraus, Northwestern University, Frances Searle Building, 2240 Campus Drive, Evanston, IL 60208, USA.
Tel: +1 847 491–3165; Fax: +1 847 491–2523;
E-mail: nkraus@northwestern.edu

effective in discerning populations with auditory-based learning problems from normal-learning controls. These differences have been identified both at the cortical and subcortical levels, and signify deficits in the pre-conscious encoding of the elemental speech signal rather than a higher-level, cognitive shortcoming.

### 2.1. Cortical response to speech-sound change (Kraus et al., 1996)

A consonant-vowel continuum, ranging from /da/ and /ga/, was synthetically constructed so that only one aspect, onset frequency of the third formant (F3), differed among its members. Ninety-one school-age children who had been clinically diagnosed with a variety of language-based learning problems (LP) and 90 normal-learning controls (NL) were tested on their ability to distinguish between closely-spaced /da-ga/ pairs. On average, LP children had much higher discrimination thresholds for stop consonants than the controls, consistent with work showing similar perceptual weaknesses in this population (Tallal and Stark, 1981; Elliott et al., 1989). To control for LP children's ability to perform the task, a continuum composed of /ba/ and /wa/, differing in the transition duration of F1 and F2, was administered, and both groups performed equivalently.

The mismatch response, or mismatch negativity (MMN), is an auditory evoked response that signals stimulus change. It is a relevant reflection of speech encoding because the speech signal itself is characterized by acoustic change. The stimulus delivery protocol for MMN recording involves a presentation of an "oddball" stimulus sequence. This consists of repeated presentation of one sound (i.e. standard), randomly replaced by another (i.e. deviant) in a small percentage of trials. Subjects are instructed to ignore the stimuli and to attend to the soundtrack of a movie in the non-test ear. If the mismatch response is present, it is seen as a negativity in the averaged response to the deviant stimulus, relative to that of the standard stimulus. A /ga-da/ pair from the same synthesized continuum used in the discrimination task was chosen as the standard and deviant stimuli

for the MMN testing protocol. The chosen contrast, differing in F3 onset by 80 Hz, was discriminable, but near discrimination threshold, for most NL children. A near-threshold /ba-wa/ pair, differing in formant transition duration by 5 ms, was chosen as a control. Forty-two children underwent the MMN testing. Half were able to distinguish between the /ga-da/ pair, half were not; all were able to discriminate the /ba-wa/ pair.

Children in both groups had robust MMNs to the /ba-wa/ pair. Conversely, in response to the /ga-da/ pair, the group comprising the good /ga-da/ perceivers had robust MMNs while the poor /ga-da/ perceivers had small or absent MMNs (Fig. 1). This indicates that the difficulties that the LP children experienced in discriminating the stop consonants is manifested in a passively elicited preconscious neural response – independent of attention and cognition – and signals a breakdown of acoustic encoding of stimulus change along the afferent auditory pathway.

### 2.2. Cortical response to rapidly-presented speech sounds in noise (Wible et al., 2002a)

Increased talker rate and background noise are two conditions known to adversely affect accurate perception and recognition of speech. Cortical responses to rapidly repeated speech sounds, both in quiet and noisy backgrounds, were investigated in LP children and NL controls.

Four-token /da/ trains were presented monaurally to the right ear at 80 dB. The stimulus was 40 ms in duration and was presented with an interstimulus interval of 360 ms, and an inter-train interval of 1060 ms. A continuous white noise masker, at a signal-to-noise ratio of +15 dB, was added to half of the trials. Averaged responses to the first and last stimulus in a train were compared, both when presented in quiet and with the masker. Inter-response correlations, which described relative changes in morphology between responses to stimuli in position 1 and position 4, and thus timing, were measured.

Under the combined stresses of repetition and noise, LP subjects demonstrated poorer inter-response correlations in noise than in quiet (Fig. 2).
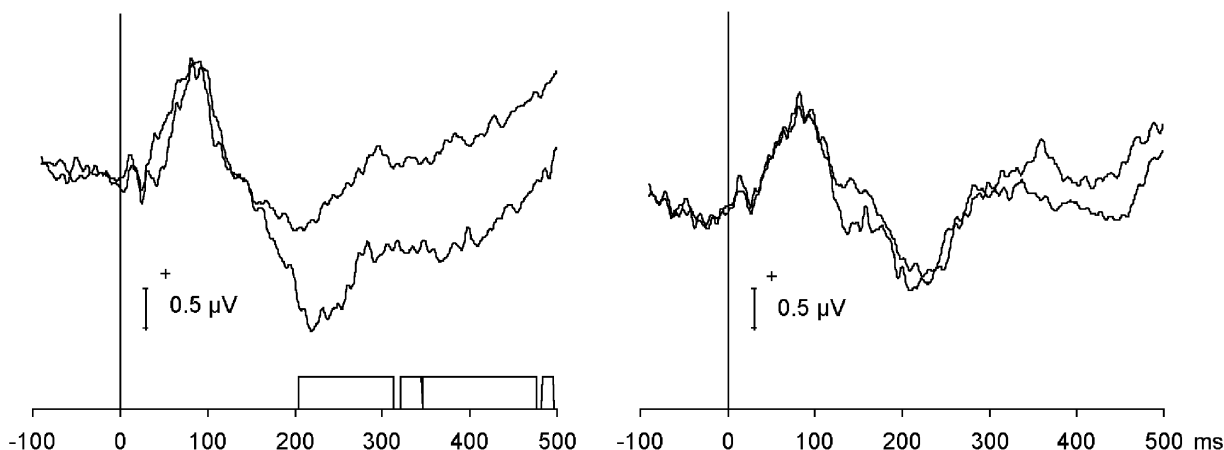
Fig. 1. Mismatch response to a speech syllable pair /ga-da/. In subjects who perceived the difference between the syllables (left), the deviant response differed significantly from the standard response from about 200 ms onward. Boxes along abscissa represent region where waveforms significantly differed ($p<0.05$). Children who were unable to perceive the differences did not exhibit a mismatch response (right). Modified from Kraus et al., 1996.

NL controls demonstrated no differences between quiet and noise on this measure of correlation between repeated responses. Such an accurate manifestation of stimulus timing is a hallmark of the normal perceptual system. Poor correlation among LP children indicates that response morphology was
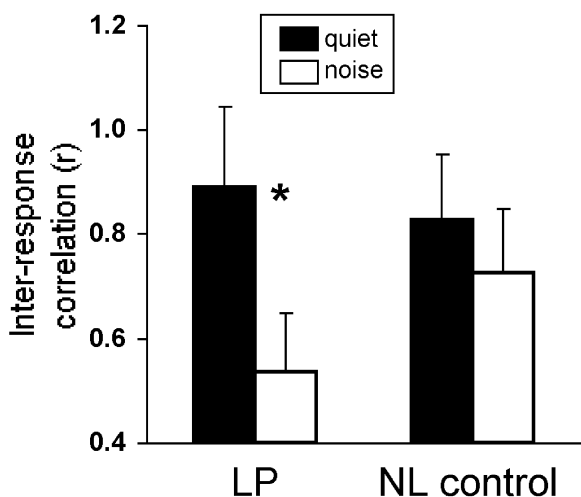


Fig. 2. Cortical inter-response correlations. In quiet, inter-response correlations are about the same for LP children (left) and normal controls (right). In background noise, however, the inter-response correlation is much poorer for the LPs. Modified from Wible et al., 2002a.

not maintained to rapidly presented stimuli in noise, which could implicate inconsistency in the timing of response generators. Across groups, the inter-response correlations bore a significant positive relationship to a standardized measure of auditory processing. These results suggests that the speech-sound perception difficulties seen in LP children may be due to degraded cortical temporal processing – the auditory system's ability to respond precisely under conditions of rapid temporal stimulation – in challenging listening conditions.

### 2.3. Subcortical response to speech sounds
(Cunningham et al., 2001; King et al., 2002; Wible et al., 2002b)

Short-latency – up to 10 ms – scalp recorded auditory evoked responses have long been used to assess hearing sensitivity and auditory pathway integrity. The auditory brainstem response to simple stimuli such as clicks and tone pips consists of a series of well-characterized peaks that reflect neural responses originating from the eighth nerve to the posterior midbrain. A novel line of research focuses on activity occurring over a similar short latency in response to speech sounds. This speech-evoked response, containing transient and sustained components, mimics acoustic aspects of speech itself.
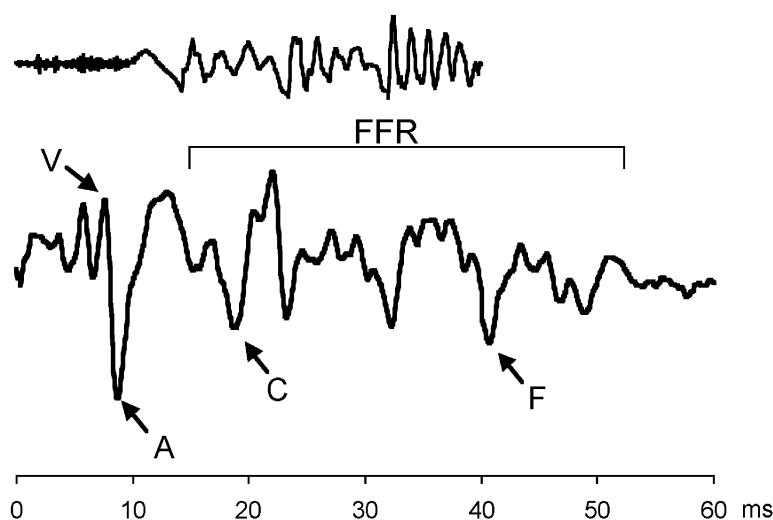
Fig. 3. Stimulus /da/ (top) and its subcortical response (bottom). The response to /da/ includes both transient and sustained components. The most stable transients, V, A, C and F, are labeled.

Inasmuch as it may be an oversimplification, there are certain parallels between consonants and vowels, and transient and sustained evoked responses.

A 40 ms syllable /da/ was presented to children with auditory-based learning problems and normal controls. Stimuli were presented both in quiet and with a continuous background white noise masker. The short-latency evoked response to this complex sound comprises a series of transient onset peaks – much like to a click or tone pip – and a sustained frequency-following response (FFR), which is phase-locked to the fundamental frequency of the speech stimulus (Fig. 3).

In addition to conventional latency and amplitude measurements of the four most stable discrete peaks, V, A, C and F, a series of analysis techniques was devised to describe the longer-lasting FFR as a whole. A broad measure of activation was measured by RMS amplitude. A more precise measure of magnitude was the amplitude of the specific frequency content in the response corresponding to the fundamental frequency (F0) and the first formant (F1) of the /da/ stimulus. Precision of timing was assessed by stimulus-to-response correlation and inter-response correlation between quiet and noise conditions (Warrier et al., 2002).

There were differences between NLs and LPs on several measures. For LPs, in quiet, response latencies of the four major peaks were significantly later. The amplitude of the F1 component of the response was suppressed. In noise, the earlier onset peaks were more frequently eliminated in the LPs. Inter-response correlations of the FFR were poorer and the amplitude of the F0 component was reduced. A significant relationship was found between the subcortical response's F1 amplitude and reading, as well as the /da-ga/ speech discrimination task described in subsection 2.1, above. Taken together, these findings reveal that normal speech perception depends on accurate encoding of sound structure – particularly the precision of responses timing – in the auditory brainstem and cortex. Moreover, the speech-sound perception difficulties in LP children may be due, in part, to degraded temporal processing under conditions that remain relatively unchallenging to normal listeners.

## 3. Improving speech-sound encoding

In recent years, there has been much interest in training programs designed to improve language skills in children with auditory-based learning problems (Tallal et al., 1996; Morrison, 1998; Diehl,

1999). Such programs may have utility for other populations, as well. Auditory training may be beneficial for older people whose speech perception has diminished and may serve as an aid in foreign language acquisition in the normal population. These programs involve intensive exposure to speech sounds; first using exaggerated cues and then gradually moving toward the subtler distinctions that are experienced in natural speech. We have been examining the underlying changes in physiology using two complementary designs. First, in a group of LP children and normal controls, the physiological and behavioral effects of cue-enhanced speech were examined. Second, a battery of physiological and behavioral tests was applied to LP children before entering a commercial training program, and then again following completion to determine whether physiological changes associated with more precise encoding of sound structure accompanied behavioral improvements.

### 3.1. Improving speech-sound encoding: change the signal (Cunningham et al., 2001)

In subsection 2.1, above, it was noted that LP children have difficulty discriminating fine-grained differences between speech syllables. Not surprisingly, this deficit is exacerbated by background noise. Background noise also has been demonstrated to more severely degrade the cortical P2/N2 evoked response to a speech sound in the LP population. In order to establish the degree to which cue-enhancements to the speech signal improve behavioral discrimination and physiology, a study examining the effects of cue-enhancement was designed.

A synthetic /ada/ to /aga/ continuum, with members differing only in the frequency of the consonant's F3 onset, was constructed. Two cue-enhancement strategies were employed – lengthening the duration of the stop gap between the initial /a/ and the /da/, and increasing the release burst intensity of the consonant /d/. Both manipulations are used naturally by speakers when attempting to speak clearly (Picheny et al., 1986). Three

additional continua were created, using each of those strategies separately and in combination. Subjects' discrimination thresholds, in quiet and with background noise, were established for all four continua. Cortical P2/N2 amplitudes also were measured to "conversational" and "clear" stimuli in quiet and with background noise.

In quiet, both NLs and LPs had equivalent discrimination scores and P2/N2 amplitudes to the conversational stimulus. With the addition of background noise, LPs' discrimination scores suffered and their cortical response amplitudes were diminished compared to the controls. In noise, the cue-enhanced clear speech stimuli restored LPs' discrimination ability and their P2/N2 response amplitudes to the same level as the controls (Fig. 4).

Thus the changes in speaking style that people naturally make when speaking to, for example, hearing impaired individuals or non-native speakers, have been demonstrated to effect a change in the cortical encoding of sound structure. Would analogous response changes occur in response to the *same* speech stimulus after an individual undergoes training to improve speech perception?

### 3.2. Improving speech-sound encoding: Change the response (King et al., 2002; Hayes et al., 2003)

Twenty-seven children with learning disabilities who were enrolled in an independently directed commercial auditory training program were subjects. Prior to enrolling in the program, and again, within three months after its completion, a battery of behavioral speech-perception tests, standardized measures of learning and academic achievement, and cortical and brainstem evoked responses was administered to them. Fifteen controls underwent the battery twice within a similar time span but received no directed training.

Several changes were seen between the pre- and post-training behavioral tasks and the physiological measures, and there were some noteworthy relationships between them. Furthermore, one pre-training
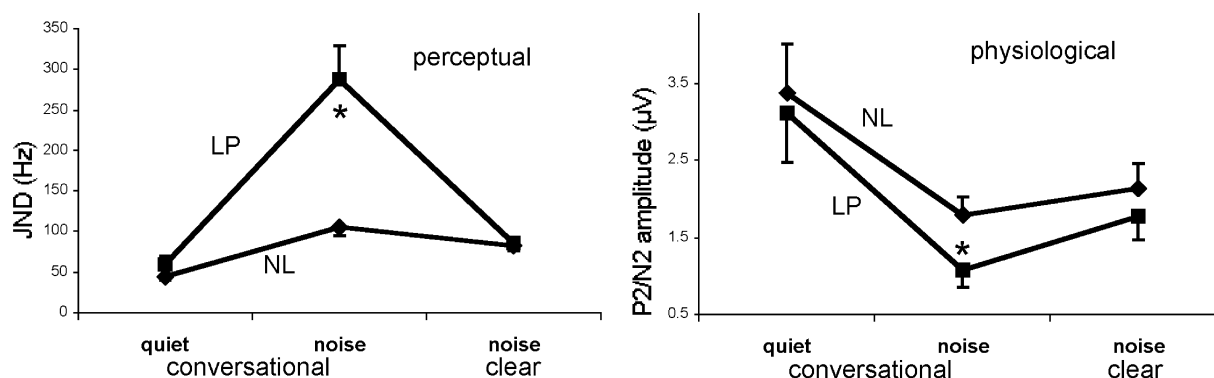
Fig. 4. Perceptual (left) and physiological (right) effects of cue-enhanced speech. Background noise affects LPs discrimination and cortical responses more severely than the controls. Cue-enhanced speech stimuli restore discrimination and physiology to normal levels. Modified from Cunningham et al., 2001.

physiology measure was predictive of behavioral gains following training.

The experimental group demonstrated significant gains on two standardized tests that assess auditory processing. In addition, changes were seen in the cortical response to /da/ in quiet. There is a normal maturational time course in both latency and amplitude of cortical evoked responses P1 and N2 (Oades et al., 1997; Sharma et al., 1997; Cunningham et al., 2000). Changes in these responses, consistent with normal maturation, were demonstrated over the short testing interval in the experimental group but not in the controls. Furthermore, the cortical response's resilience to background noise, as measured by inter-response correlation, improved following training only in the experimental group. The subjects whose subcortical response latencies were delayed were those who showed the most gains: cortical responses in noise most improved in resilience and /da-ga/ discrimination improved. Finally, there is some evidence that brainstem encoding itself may change following training in some subjects.

Thus, a variety of physiological indicators either accompanied or predicted behavioral gains following commercial auditory training. The physiological data indicate that the preconscious encoding of sound structure is plastic; i.e. it is not hard-wired but is modifiable by learning and experience. Such preconscious, non-cognitive, physiological tests may be valuable in predicting which children may receive the most benefit from training programs and serve as tools for monitoring their progress.

## 4. Conclusions

Accurate speech-sound encoding requires an auditory pathway that maintains the precise timing features that compose speech; the phasic and tonic aspects of ensemble neural firing share many features of the speech signal itself. The transient responses recorded from the brainstem and cortex and the sustained responses originating in the midbrain in combination can tell us a great deal about the integrity of the speech-sound encoding mechanism. A number of physiological abnormalities – neural timing – have been identified in a population of children who experience auditory – based learning problems – some linked to specific behavioral deficits.

Importantly, there is evidence that the neural encoding of elemental acoustic events can be altered. Cue-enhanced stimuli, themselves, are useful in effecting a normal-like response in an LP individual, and their use as a training tool can lead to improved neural timing to non-enhanced speech. The malleability of encoding of acoustic sound structure in the auditory pathway suggests approaches that could be applied more generally in other instances where improved perception of sound is desirable, such as learning music or foreign languages.

630

## References

Cunningham, J., Nicol, T., Zecker, S. and Kraus, N. Speech-evoked neurophysiologic responses in children with learning problems: development and behavioral correlates of perception. *Ear and Hearing*, 2000, 21: 554–568.

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

Diehl, S. Listen and Learn? A software approach review of Earobics. *Language, Speech and Hearing Services in Schools*, 1999, 30, 108–116.

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

Hayes, E., Warrier, C.M., Nicol, T., Zecker, S.G. and Kraus, N. Neural plasticity following auditory training in children with learning problems. *Clin. Neurophysiol.*, 2003, 114: 673–684.

King, C., Warrier, C.M., Hayes, E. and Kraus, N. Deficits in auditory brainstem encoding of speech sounds in children with learning problems. *Neurosci. Lett.*, 2002, 319: 111–115.

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

Morrison, S. Computer applications: Earobics Pro. *Child Language Teaching and Therapy*, 1998, 14: 279–284.

Oades, R.D., Dittmann-Balcar, A. and Zerbin, D. Development and topography of auditory event-related potientials (ERPs): mismatch and processing negativity in individuals 8–22 years of age. *Psychophysiology*, 1997, 34: 677–693.

Picheny, M.A., Durlach, N.I. and Braida, L.D. Speaking clearly for the hard of hearing. II: Acoustic characteristics of clear and conversational speech. *J. Speech and Hearing Res.*, 1986, 29: 434–446.

Sharma, A., Kraus, N., McGee, T.J. and Nicol, T.G. Developmental changes in P1 and N1 central auditory responses elicited by consonant-vowel syllables. *Electroencephalogr. Clin. Neurophysiol.*, 1997, 104: 540–545.

Tallal, P. and Stark, R.E. Speech acoustic-cue discrimination abilities of normally developing and language-impaired children. *J. Acoust. Soc. America*, 1981, 69: 568–574.

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

Warrier, C.M., Johnson, K.L., Nicol, T.G. and Kraus, N. Learning-impaired children exhibit timing deficits in cortical responses to speech stimuli presented in background noise. *Ass. Res. Otolaryngol.*, 2002, abstract.

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

Wible, B., Nicol, T. and Kraus, N. Effects of stimulus repetition and background noise on speech evoked auditory brainstem and frequency following responses in normal and learning impaired children. *Ass. Res. Otolaryngol.*, 2002b, abstract.