Published in final edited form as:

Cortex. 2005 June ; 41(3): 327–341.

INDIVIDUAL DIFFERENCES IN AUDITORY PROCESSING IN SPECIFIC LANGUAGE IMPAIRMENT: A FOLLOW-UP STUDY USING EVENT-RELATED POTENTIALS AND BEHAVIOURAL THRESHOLDS

Dorothy V.M. Bishop and Genevieve M. McArthur

(University of Oxford, UK)

Abstract

It has frequently been claimed that children with specific language impairment (SLI) have impaired auditory perception, but there is much controversy about the role of such deficits in causing their language problems, and it has been difficult to establish solid, replicable findings in this area. Discrepancies in this field may arise because (a) a focus on mean results obscures the heterogeneity in the population and (b) insufficient attention has been paid to maturational aspects of auditory processing. We conducted a study of 16 young people with specific language impairment (SLI) and 16 control participants, 24 of whom had had auditory event-related potentials (ERPs) and frequency discrimination thresholds assessed 18 months previously. When originally assessed, around one third of the listeners with SLI had poor behavioural frequency discrimination thresholds, and these tended to be the younger participants. However, most of the SLI group had age-inappropriate late components of the auditory ERP, regardless of their frequency discrimination. At follow-up, the behavioural thresholds of those with poor frequency discrimination improved, though some remained outside the control range. At follow-up, ERPs for many of the individuals in the SLI group were still not age-appropriate. In several cases, waveforms of individuals in the SLI group resembled those of younger typically-developing children, though in other cases the waveform was unlike that of control cases at any age. Electrophysiological methods may reveal underlying immaturity or other abnormality of auditory processing even when behavioural thresholds look normal. This study emphasises the variability seen in SLI, and the importance of studying individual cases rather than focusing on group means.

Keywords

specific language impairment; auditory; ERP; maturation

Introduction

The failure of some children to master language at the normal rate, despite good intelligence and adequate instruction, has puzzled researchers for several decades. One theoretical account proposes that specific language impairment (SLI) in children is caused by low-level auditory perceptual problems. Most children with SLI have literacy problems, and the theory has also been extended to account for developmental dyslexia.

Early studies of auditory perceptual deficits in children with SLI and dyslexia were conducted using the Auditory Repetition Test (ART, see Tallal and Piercy, 1978, for review). The ART requires a listener to press buttons to correspond to a sequence of brief tones that vary in frequency. Children with SLI performed poorly on this task if the tones were presented with an interstimulus interval (ISI) of less than around 250 ms, but they performed accurately at longer ISIs. This led Tallal and colleagues to propose that SLI results when the auditory system responds sluggishly. This will affect language development, because when listening to speech, processing of one sound may not be complete before the arrival of the next sound, and so formation of phonemic categories will be disrupted. Although several subsequent studies have found results consistent with this “temporal processing” hypothesis (see Habib, 2000; Wright et al., 2000, for reviews), others have failed to find the predicted perceptual deficits in people with dyslexia or SLI, both on the ART (Bishop et al., 1999a) and on other auditory temporal measures such as backward masking (Bishop et al., 1999b) or gap detection (McAnally and Stein, 1996; see also Rosen, 2003, for review).

Evidence for Spectral Deficits in SLI and Dyslexia

One issue that has arisen is whether poor performance on the ART may stem from deficiencies in spectral rather than temporal processing of auditory signals. A person with a spectral difficulty would have problems distinguishing sounds that differ in frequency, regardless of their duration or presentation rate. Evidence for poor frequency discrimination is apparent in some studies using the ART. First, many studies find a subset of children have to be excluded because they cannot learn the initial discrimination between tones of different frequency, even under optimal conditions when there is no time pressure (Bishop et al., 1999a; Breier et al., 2002; Heath et al., 1999; Reed, 1989; Tallal et al., 1981). Further, several studies have found that many children who do learn the initial tone discrimination to criterion have problems distinguishing tone sequences regardless of whether they are presented at slow or fast rates (Bishop et al., 1999a; Lincoln et al., 1992; Marshall et al., 2001; Waber et al., 2001). In a similar vein, Baldeweg et al. (1999) found that dyslexics were inferior to controls at distinguishing a 50-ms, 1000-Hz tone from deviant tones that ranged from 1015 to 1090 Hz. In addition, several studies have found usually high frequency difference limens (i.e. the smallest difference in frequency that can be detected) in people with language or literacy problems (Ahissar et al., 2000; Cacace et al., 2000; Hari et al., 1999; McAnally and Stein, 1996).

Further evidence of anomalous spectral processing in SLI was reported by Wright et al. (1997). They showed that poor backward masking performance of children with SLI could be ameliorated by placing a wide spectral ‘notch’ in the masking noise (i.e. using noise that excluded frequencies at or near the signal frequency). Children with SLI could detect a tone immediately followed by noise relatively well, provided the tone and noise were sufficiently different in frequency.

Results such as these raise the possibility that the principal reason for deficient performance on the ART is poor frequency discrimination, and that this is most evident when the task is made difficult by using a rapid presentation rate.

Variability within the Learning Disabled Population

Although there is evidence of frequency discrimination deficits in people with language and literacy problems, it is clear that not every individual is impaired. As stressed by Hill et al. (1999) and Amitay et al. (2003), there is substantial heterogeneity among poor readers on tests of frequency discrimination, with some doing very poorly and others scoring well within normal limits. McArthur and Bishop (2003) compared 16 young people with SLI and

16 controls on tests of frequency discrimination and recognition backward masking, and found that only around one third of the SLI group had evidence of auditory deficits, and these were poor at the basic frequency discrimination task, rather than showing selective problems on the backward masking task. Clearly, if auditory deficits are seen in only a subset of individuals, then one may mask genuine group differences by combining heterogeneous cases.

Maturational Considerations

As well as being aware of individual differences in auditory deficit, it is important to take into account the protracted course of auditory development in humans. Although there is ample evidence that the cochlea and auditory brainstem pathways are adult-like in young infants (Ponton et al., 2000), neuroanatomical studies by Moore (2002) have demonstrated striking changes in auditory cortex occurring throughout childhood and early adolescence. Mature axons in cortical layers 2 and 3, which are involved in cortico-cortical connections, do not begin to appear until 5 years of age and only reach adult levels by 11 or 12 years of age. As Moore (2002) pointed out, this late maturation of the auditory cortex could explain why certain aspects of auditory function do not reach adult levels until early adolescence: In general, adult levels of performance are achieved only by older children or adolescents on tasks that involve cortical processing, rather than more peripheral auditory mechanisms. These include perception of distorted or degraded speech, detection of a backward masked signal, and binaural unmasking, which assesses the ability to perceive small differences in the phase of auditory signals presented to the two ears (Hall and Grose, 1990; Hartley et al., 2000; Hogan and Moore, 2003).

Most auditory deficit accounts of SLI implicitly assume that language development is 'deviant' in affected individuals, because some crucial perceptual skill is chronically deficient. However, the literature on auditory development reviewed above suggests another possibility, which we shall refer to as the 'maturational hypothesis'. This maintains that any deficits seen in children with SLI or dyslexia reflect delayed maturation of cortical development, rather than a more permanent abnormality. This raises the question of whether auditory deficits are more commonly found in younger rather than older participants with SLI or dyslexia. McArthur and Bishop (2001) reviewed the relevant literature and found that this was not the case. Indeed, some of the most robust evidence for auditory deficits has been found in adult samples. However, our survey included a variety of auditory tasks that may have different developmental trajectories. To evaluate the maturational hypothesis adequately, one needs to see the same people assessed on the same task at different points in development. Bernstein and Stark (1985) adopted this approach and found that performance on the ART did change, and sometimes normalise, as children with SLI grew older. Wright and Zecker (2004) found that backward masked thresholds were impaired in 8-year-olds with learning disabilities but were normal in 12-year-olds. However, the older children did more poorly than controls on a simultaneous masking task. They suggested that the changing profile of auditory deficit might be explained in terms of auditory system immaturity. Similar arguments have been advanced by Hautus et al. (2003), who used a gap-detection task, and showed that whereas 6- to 9 year-olds with dyslexia were significantly impaired relative to controls, 10-to 13-year-olds were not. In this regard, it is noteworthy that the subset of people with poor frequency discrimination in the sample of McArthur and Bishop (2004) tended to be younger than the SLI cases with normal frequency discrimination thresholds.

The maturational hypothesis is of particular interest because it suggests that in focusing on whether any deficit is spectral or temporal we may be missing the point. If a person with SLI has auditory processing skills equivalent to a typically developing person some 3 or 4 years younger, then whether or not a deficit is seen on a given task will depend on normal

developmental trajectory of performance on that task. As Wright and Zecker (2004) suggest, the same individual could show deficits on temporal or spectral tasks at different points in development. For instance, we know from studies such as that of Thompson et al. (1999) that frequency discrimination is a relatively late-maturing auditory skill, with typically developing children showing a sharp improvement in thresholds from around 6 to 9 years, after which adult levels of performance are attained. Suppose that a typical child with SLI performs like a typically-developing child some 3 or 4 years younger. In this case, we would continue to see poor frequency discrimination in children with SLI up to 12 or 13 years of age, after which they would appear to catch up with their peers.

The maturational hypothesis may seem implausible because it seems to entail that a child with SLI should eventually catch up with the peer group in auditory and language skills, whereas SLI typically persists in adulthood. However, there are two points to consider. First, any person whose language skills lagged three or four years behind age level in childhood would have an enormous amount of ground to make up because of the serious educational impact of limited language and literacy skills. Second, as Wright and Zecker (2004) have suggested, it is possible that limits on maturation are imposed by puberty, in which case SLI would be characterised both by an initial delay in childhood followed by a plateau when a person reaches adolescence.

ERP Indicators of Auditory Processing

To date, evidence in support of the maturational hypothesis has come from behavioural studies. In this paper we suggest that electrophysiological data can provide a complementary source of evidence. Figure 1 shows a typical adult ERP elicited by a brief auditory stimulus. This is obtained by amplifying and then averaging the electrical activity from electrodes situated on the scalp, as the participant hears the same stimulus repeatedly. The portion of the waveform shown in Figure 1, extending from 50 ms to around 500 ms post stimulus onset is often known as the ‘late’ auditory ERP, and reflects cortical activity. A characteristic profile of peaks and troughs is typically observed. These components of the waveform are labelled by polarity and order of appearance, i.e., P1, N1, P2, and N2 are the first positive, first negative, second positive, and second negative peaks respectively.

Although there are some relationships between the physical characteristics of an auditory stimulus and the latency or size of the components observed, an averaged waveform such as that in Figure 1 cannot on its own tell us how well a person discriminates sounds. Furthermore, ERPs do not provide precise information about the localisation of underlying brain events, though statistical methods such as dipole source modelling have been developed that allow one to infer the source of an observed component if activity is measured across a large array of electrodes (Scherg and von Cramon, 1986). To complicate matters further, late auditory ERPs are known to reflect the combined activity of several underlying brain regions. For example, N1 may result from activity in the primary auditory cortex, the postero-superior temporal plane, and non-specific frontal areas (Bruneau and Gomot, 1998); P2 might be generated by two areas in the supratemporal auditory cortex (Näätänen, 1992 ; Tonnquist-Uhlen, 1996a) or in the mesencephalic reticular activating system (Ponton et al., 2000); and N2 may originate in the superior temporal gyrus and medial temporal lobes (O'Donnell et al., 1993).

Perhaps the simplest question one can ask of an ERP is whether the peaks and troughs in the waveform occur at the expected latencies after stimulus onset. One way of interpreting the auditory temporal processing account of SLI would be to suppose that the responses of the central auditory system to sound might be sluggish, predicting that the components of the auditory ERP should occur later than usual. A handful of studies have compared auditory ERPs for individuals with SLI or dyslexia and a control group, but results have been

variable. Further, where differences have been found, these usually affect the amplitude rather than the latency of particular components. Adams et al. (1987) found unusually large P1 in children with SLI, whereas Pinkerton et al. (1989) found that N1 to tones was smaller in children with literacy problems than in a matched control group. However, these findings were not replicated for children with severe SLI by Mason and Mellor (1984), Neville et al. (1993), Marler et al. (2002) or Ors et al. (2002), nor for adult dyslexics studied by Baldeweg et al. (1999), Helenius et al. (2002), Nagarajan et al. (1999), or Renvall and Hari (2002). Lincoln et al. (1995) found *larger* N1 in language-impaired children in some conditions of their study. One study that did report increased latencies of ERP components was that by Tonnquist-Uhlén (1996b), who studied N1, P2 and N2 in 20 children with severe language impairment and reported that latencies of all components were significantly longer than for control children. However, standardized test results on IQ or language status were not available for this sample, and a high proportion had abnormalities of the brain-stem evoked response and/or pathological electroencephalograms (EEGs), suggesting they would not meet usual criteria for SLI. Another study, by Byring and Järvillehto (1985), reported increased latency of P1 for 13-year-olds with spelling difficulties compared with controls. However, this was not replicated by Pinkerton et al. (1989).

Other studies have measured the mismatch negativity (MMN) in people with language or literacy problems. This ERP component is elicited using an 'oddball' design, in which the listener is presented with a sequence of standard stimuli, with occasional 'deviant' stimuli occurring on a minority of trials. The brain response to the deviant stimuli is averaged and subtracted from the average response to standards. For discriminable stimuli, the deviant typically elicits an additional negativity around 200 ms after stimulus onset. Näätänen and Alho (1995) reviewed work on the MMN and concluded that the MMN reflects a mismatch in sensory memory between the representation of the standard stimulus and the deviant stimulus. Thus the MMN indexes the brain's ability to distinguish between, rather than simply to detect auditory stimuli. Several investigators have reported a diminished MMN to deviants that differed in frequency from a standard tone (Baldeweg et al., 1999; Holopainen et al., 1997, 1998; Korpilahti 1995; Korpilahti and Lang, 1994), but others have failed to find this effect (Schulte-Körne et al., 1998; Uwer et al., 2002).

Overall, then, the findings from studies using ERPs with cases of language or literacy impairment give results that are at least as inconsistent as those from behavioural studies. Inconsistent findings with the MMN could reflect its relatively low reliability at the individual level, but this would not seem an adequate explanation for the variable results obtained with other late components of the ERP (i.e., P1, N1, P2 and N2; Escera and Grau, 1996; Kathmann et al., 1999; McArthur et al., 2003; Uwer and von Suchodoletz, 2000). Just as with the behavioural studies, it seems necessary to take into account heterogeneity within the SLI population, when trying to make sense of such data.

Using ERPs to Study Individual Differences in SLI

In an attempt to do this, McArthur and Bishop (2004) evaluated how closely the individual ERPs of people with SLI matched the average waveform of controls of similar age in the N1-P2-N2 region. (The method is described more fully below). We had expected to find abnormal ERPs in the subset of SLI cases who had poor frequency discrimination, but instead, we found that the majority of SLI cases had age-inappropriate ERP waveforms, regardless of whether or not they had normal thresholds on a frequency discrimination task. However, the waveforms of individuals with SLI differed markedly from one case to the next, emphasising the importance of analysing results at the individual level.

Using ERPs to Study a Maturational Explanation of SLI

A further way in which ERPs may inform our understanding of SLI is by providing a neurobiological measure of maturity of auditory processing. Recent large-scale studies documenting developmental changes in ERPs throughout childhood have shown striking changes from childhood to adolescence. A study of 108 individuals aged from 5 years to adulthood was conducted by Albrecht, von Suchodoletz and Uwer (2000) using 1000 Hz tones to elicit ERPs. They seldom observed N1 and P2 in children below the age of 13 years, but these components emerged in the course of adolescence. The location of dipole sources of the ERP did not change during development, but the relative activity of tangential and radial dipole sources changed substantially, and the N1-P2 complex became evident in adolescence in a dipole that reflected activity in the supra-temporal plane. Ponton et al. (2000) measured ERPs at 30 scalp locations to brief click trains in 118 participants aged from 5 to 20 years. The N1 and P2 components were not evident in children aged around 7 to 8 years, but gradually emerged around 9 years of age and become more pronounced with age, assuming an adult-like form by 15–16 years. Furthermore, the latency and amplitude of P1 diminished with age. Similar findings were reported by Sharma et al. (1997) who investigated ERP to consonant-vowel stimuli in 86 children aged from 6 to 15 years. N1 was seen in around two thirds of those aged below 13 years, but in all children aged 13 years and above. Other studies have shown that the likelihood of observing the N1-P2 complex in children depends not only on age but also on methodological issues, such as the specific stimuli used and the interval between stimuli (Ceponiene et al., 1998; Karhu et al., 1997). The recording site is also important: Several studies have shown that N1 is larger at temporal sites in children, and at the vertex in adults (e.g., Pang and Taylor, 2000). Daruna and Rau (1987) suggested that N1 does not appear at the vertex until age 7 years. The fact that such clear maturational changes can be seen on the ERP suggests that electrophysiological data could be used to test the idea of a ‘maturational lag’, by seeing how far the ERPs of older individuals with SLI resembled those of younger typically developing children.

Aims of the Current Study

In the current study we re-tested a subset of the original group of children and young people seen by McArthur and Bishop (2004), plus some additional cases, 18 months after the initial study. The study had three main aims:

1. To test whether individuals who had poor frequency discrimination (FD) thresholds when originally seen by McArthur and Bishop (2004) at Time 1 would show comparable deficits some 18 months later, or whether, as predicted by a maturational account, their thresholds would improve with age.
2. To replicate the original finding of McArthur and Bishop (2004) that individual waveforms of people with SLI deviate from the grand average for controls in their age range, in the time window corresponding to N1-P2-N2.
3. To consider whether the age-inappropriate waveforms of individuals in the SLI group resemble waveforms of younger typically-developing children, as would be predicted by a maturational account, or whether they are deviant, i.e. not resembling control cases at any age.

In the current study, we initially hoped to study the MMN in individual participants, but our data were too unreliable to give meaningful results at the individual level (cf. Taylor and Baldeweg, 2002; Uwer and von Suchodoletz, 2000), and so this aspect of the ERP will not be considered further.

Methods

Participants

We followed up 24 of the original cases seen by McArthur and Bishop (2004): 11 with SLI and 13 controls. The SLI group included four individuals with poor-FD (i.e., threshold for distinguishing a comparison tone from a 600 Hz standard was greater than 700 Hz) when first assessed, six with good-FD, and one for whom data were not available because of equipment failure. We shall refer to the initial data from the McArthur and Bishop study as Time 1, and the new data obtained for the current study as Time 2. To increase power of Time 2 computations, an additional five cases with SLI and three controls were recruited from the same sources as the original sample (i.e., language development centres and support groups for the SLI sample, and scout and guide groups, a college, and a high school in Oxford for the controls). However, one of the new controls had to be excluded because of excessive artefact in his ERP (437 epochs accepted, compared with between 1091 and 1395 epochs for other participants), so the final sample consisted of 16 cases with SLI (mean age 15.82 years, SD 2.70 years, range 12 to 20 years) and 15 controls (mean age 15.18 years, SD 2.58 years, range 12 to 21 years). One of the new recruits with SLI had a FD threshold greater than 700 Hz. The rest of the new recruits had FD thresholds below 700 Hz.

Diagnostic Tests

Diagnostic tests that had been administered at Time 1 were given to the new recruits for this study. To be included in the SLI group the participant had to score more than one standard deviation below the age-appropriate level on at least two of four standardised tests of spoken language (see Table I). To be included in the control group, participants had to score no lower than one standard deviation below the mean on at least three of the four spoken language tests. Participants in both groups had to have non-verbal scaled score of at least 75, and pass a audiological screening test to ensure that the threshold for detecting a 750 Hz tone was no higher than 20 dB. The SLI and control groups were well matched on both age and non-verbal ability at Time 1 (McArthur and Bishop, 2004) and Time 2 (see Table I).

Frequency Discrimination Thresholds

The frequency discrimination task used at Time 1 and 2 comprised 10 practice trials and up to 60 experimental trials. Each trial was composed of two tones that were separated by 500 ms silence. Each tone was visually represented on the PC monitor by a square button ("1" for the first tone and "2" for the second tone) that flashed when the tone was played. One tone was 80 dB SPL in intensity, 25 ms (including 2.5 ms onset and offset) in duration, and had a frequency of 600 Hz (the standard). The other tone was the same except that it had a higher frequency. Tones were presented through headphones, and the participant's task was to identify which interval (1 or 2) contained the higher tone. If a response was correct, a coloured "thumbs-up" sign was presented on the monitor. An incorrect response triggered a plain black cross.

The frequency of the higher tone was initially set at 700 Hz (the ceiling value was 800 Hz), and was adjusted between trials in 25 Hz steps using a one-up, three-down adaptive procedure (Shelton and Scarrow, 1984), converging on a threshold value at which 79 per cent of responses were correct. The FD threshold was the mean frequency of the higher tone across the last even number of step-size reversals after the first four reversals in response adjustment. Higher threshold scores represented poorer frequency discrimination.

Auditory ERPs

Stimuli were divided into four blocks that were presented in random order. Each block was composed of 500 trials (Time 2 differed from Time 1 in this regard; Time 1 used blocks of

250 trials). A standard oddball paradigm was used: Two of the four blocks used a 25 ms, 80 dB SPL, 600 Hz tone as a standard stimulus (85% of trials) and a 25 ms, 80 dB SPL, 700 Hz test tone as a deviant stimulus (15% of trials). The other two blocks used the reverse. We had selected the oddball paradigm because of our initial interest in studying the MMN, but for the current report, we describe data from standards only.

Participants were seated in a lounge chair in an electrically shielded testing booth. The tones were presented diotically through Sennheiser HD265 headphones whilst the participant watched a video on a small television 1.3 m away. At Time 1, the mean interval between the start of each tone was 1084 ms, randomly jittered for each trial with SD of 99 ms. At Time 2, to reduce the length of the test session, we used a mean interval between the start of each tone of 734 ms, with SD 78 ms. The soundtrack of the video was played at a low-level (approximately 50 dB SPL) to divert attention from the tones. (We have subsequently done a study showing that the video soundtrack disrupts the MMN, but has much less effect on the other late ERP components: see McArthur et al., 2003).

The EEG was recorded from non-polarised sintered electrodes positioned according to the 10–20 International system. Eight electrode sites were used at both Time 1 and Time 2: two midline sites (Fz, FCz,) and three sites over each hemisphere (F3/F4, F7/F8, FC3/FC4). The ground electrode was positioned on the midline between FPz and Fz. Linked mastoids were used for the online reference. The signal was amplified 20,000 times, sampled at 250 Hz, and band-pass filtered on-line at .05–30 Hz.

The EEG was processed offline. We used standard artefact reduction algorithms to reject trials contaminated by eye movements, which were recorded from the vertical electro-oculogram (VEOG) and horizontal electro-oculogram (HEOG). The EEG was then cut into 530-ms epochs with a 50 ms pre-stimulus interval and these were baseline-corrected from –50 to 0 ms. Epochs with changes in HEOG or EEG activity greater than 150 mV from baseline were rejected. The auditory ERP of each participant was calculated by averaging all the epochs of the standard stimuli (600 Hz and 700 Hz) except for those presented immediately after a deviant. Consistent with McArthur and Bishop (2004), auditory ERPs were represented by activity at Fz because it recorded the largest response, it is the site used most often to represent auditory ERPs, and analogous N1 and P2 responses can be measured at Fz in adults and children (Ponton et al., 2000). In addition, to ensure that any group differences that were found were not specific to this one electrode, average waveforms from the Fz, FCz, F3, F4, FC3 and FC4 were also computed and compared between groups (F7 and F8 were not included in this average because activity at these sites was influenced by VEOG activity in several participants).

Analytic Approach

The usual approach adopted for analysing ERP data is to compare groups in terms of the amplitude and latency of P1, N1 and P2, measured at several different electrode sites. We did not adopt this approach for three reasons. First, as noted by Leppänen and Lyytinen (1997), the more comparisons carried out, the greater the likelihood of making a Type II error (i.e., obtaining a spurious difference by chance). A second problem is that some ERP components may not be clearly present in children's data. Most peak detection software will provide a value for a peak in such a situation, because the algorithm works by simply finding a minimum or maximum in a specified time window. So, for instance, if the amplitude declines throughout an interval that is supposed to contain the N1, a negative 'peak' will be identified at the end of the window, when the amplitude is lowest. Inclusion of such false peaks will add noise to an analysis. The third problem we confronted was that our previous behavioural studies had indicated there was considerable heterogeneity within the SLI population. Therefore, rather than comparing group averages, we needed an index

that reflected the extent to which each individual's waveform in the N1-P2 region was age-appropriate.

The solution to these problems that was adopted by McArthur and Bishop (2004) was to compute a single index of overall similarity between each individual's waveform and the grand average waveform of a control group of the same age range. This avoids the problem of inflated p values that arises when multiple statistical tests are conducted, and does not require that particular peaks be present in the waveform. The statistic that we used to measure similarity was the intra-class correlation (ICC), which is similar to the Pearson product-moment correlation, except that it takes into account differences in amplitude as well as shape between two waveforms, and so is usually much lower than the Pearson r.

The ICC between two arrays¹ is computed by the formula (MS between – MS within)/(MS between + MS within), where X and Y denote the two arrays, and:

$$\text{MS between} = ((\sum (X^2) + \sum (Y^2) + 2 \sum (X.Y)) / 2 - (\sum (X) + \sum (Y))^2 / (2 * N)) / (N - 1)$$

$$\text{MW within} = (0.5 * (\sum (X^2) + \sum (Y^2)) - \sum (X.Y)) / N$$

N = number of pairs of data points

McArthur and Bishop (2003) showed that there was good agreement between the ICC and rankings of ERP similarity made by raters who were blind to group status. This provided validation of the ICC as an objective global index of correspondence between two waveforms. However, ERP data do not fulfil the statistical assumptions of the ICC, that observations should be independent, insofar as the data points in a waveform show a high degree of autocorrelation. Therefore, to estimate whether a given ICC value was statistically different from zero, random series of data points were simulated so as to have an autocorrelation profile similar to that seen in our ERP data, and ICCs computed between pairs of these datasets for 3000 iterations. (These simulations used 33 data points, to be equivalent to our ERP analysis – see below). The distribution of ICCs from these simulations gave estimates of .50, .69 and .72 as values corresponding to the 5%, 2.5% and 1% levels of significance (1-tailed).

A further question we had to deal with was how to define the normative waveform against which the SLI waveforms would be compared. One option would be to use the grand average waveform from all the controls, but this was not ideal, because there are significant developmental changes in the waveform over the age range that we studied (e.g., Albrecht et al., 2000; Ponton et al., 2000). The ideal solution would be to compute comparison grand average waveforms for each year band separately, but we did not have adequate normative data for this to be viable. When analysing Time 1 data, McArthur and Bishop (2004) adopted a compromise solution, doing a median age split on the controls, yielding two grand average waveforms. One was based on eight controls aged below 14 years and the other was based on eight controls aged 14 years and over. Consistent with data from Albrecht et al.

¹This formula gives the same value as SPSS reliability analysis, one-way random model. We did not use Neuroscan software to compute ICCs, because it gives values that differ from those computed by this formula.

(2000), the N1-P2 complex was clearly evident in the older control group but not in the younger group.

For the analysis of Time 2 data, we adopted the same approach. Because control children were older at Time 2 than at Time 1, the numbers of control children in the two age subgroups were not equal: there were six control children aged 12.34 to 13.67 years who formed the grand average waveform for the young group, and nine young people aged 14.71 to 21.02 years who formed the grand average for the old group. Waveforms for six individuals with SLI younger than 14.5 years were compared with the young control grand average waveform, and the 10 SLI cases aged above 14.5 years were compared with the old control grand average waveform.

The interval for analysis was from 100 ms to 228 ms post stimulus onset, incorporating 33 data points at 4 ms intervals. This interval was selected because it contained the components associated with auditory stimulus analysis (i.e., the N1-P2 complex and N2; Näätänen and Winkler, 1999)².

We also needed to compute ICCs for individual *control* participants to assess how age-appropriate their waveforms were. This comparison would be biased if we compared the waveform of an individual to that of a grand mean that included their own waveform. Thus a new grand average comparison waveform was computed for each control participant, based on all the other controls of the same age, but excluding their own waveform.

To recap, the method of analysis adopted here was comparison of an individual's waveform in the range 100 to 228 ms with that of a grand average from controls of the same age range. The ICC provided a quantitative index of overall similarity of waveforms. T-tests were then used to compare the ICCs of control and SLI groups, both at Time 1 and at Time 2, after applying the Fisher z-transformation to normalise the data. The η^2 statistic was used to estimate effect size.

Results

Do FD Thresholds improve with Time?

Our first aim was to consider whether people who had poor FD at Time 1 (i.e., those with Time 1 thresholds of 700 Hz or more) would continue to show such a severe deficit at Time 2. As shown in Figure 2, there was a reasonably high correlation between the FD thresholds seen on the two occasions for the 23 participants with data at both Time 1 and Time 2; Spearman $r = .64$, $p = .001$. However, as is clear from the scatterplot, the thresholds for the SLI poor-FD group improved with time (from mean of 750 Hz to 674 Hz) whereas those for the other groups were much more stable (SLI-normal FD group had mean of 633 Hz at Time 1 and 636 Hz at Time 2, and controls had mean of 629 Hz at Time 1 and 624 Hz at Time 2). In two of the original poor-FD cases, the thresholds had moved within the limits of the control group by Time 2. Overall, then, there was evidence that the poor FD thresholds seen in some of the SLI cases at Time 1 improved over time, as would be predicted from a maturational hypothesis.

Another prediction from the maturational account is that we would not expect to see any relationship between age and FD threshold in the control group, all of whom are over the age at which adult levels of FD are achieved (estimated at around 9 years by Thompson et

²This differs from the interval of 128 to 256 ms used by McArthur and Bishop (2004). In their report, data from Time 1 were filtered with a 10 Hz high pass filter, leading to later peaks for N1 and P2. At the suggestions of reviewers, data from Time 1 and Time 2 were processed for this study with a 30 Hz low bandpass filter.

al., 1999). However, we would expect a relationship with age in the SLI group, with the younger cases performing more poorly, because at chronological age of 10 to 12 years their auditory skills would be more like a child of 6 to 9 years. There was tentative support for this at Time 1. The Spearman correlation between age and FD threshold at Time 1 was nonsignificant for the whole sample ($\rho = -.221$, $N = 23$), but participants in the SLI poor-FD group were on average younger at Time 1 than the good-FD SLI cases: poor-FD $M = 12.79$ years, $SD = 1.67$, $N = 4$; good-FD $M = 15.51$, $SD = 2.63$, $N = 10$; $t(12) = 1.89$, $p = .041$, one-tailed. At Time 2, there was a similar trend, but only three cases (two shown in Figure 2, and one of the new cases) had FD thresholds above 700 Hz, and the age difference was not significant (poor-FD $M = 14.37$ years, $SD = 1.57$, $N = 3$; good-FD $M = 16.14$, $SD = 2.84$, $N = 13$; $t(14) = 1.03$, $p = .161$, one-tailed). Thus these data are compatible with the idea that auditory skills show a maturational lag in children with SLI, with FD thresholds reaching adult levels a few years later than in typically developing children.

Do Individuals with SLI Show Age-appropriate ERPs in the N1-P2-N2 range?

Our second aim was to see whether the original finding of McArthur and Bishop (2004) could be replicated, showing that individuals with SLI are more likely than controls to have age-inappropriate ERP waveforms. As described above, the ICC was computed for the portion of the waveform from 100 to 228 ms post stimulus between each individual's waveform and the grand average of younger or older controls, depending on the age of the participant (above or below 14.5 years).

At Time 1, we found that ICCs were significantly lower for participants in the SLI group than for the control group, regardless of their behavioural frequency discrimination thresholds. After Fisher z transformation, the SLI mean was .03 ($SD = .453$) and the control mean was .54 ($SD = .491$), giving $t(30) = 3.03$, $p = .005$, $\eta^2 = .23$. At Time 2, the Fisher transformed ICC of the SLI group ($M = .19$, $SD = .37$) was again significantly lower than the transformed ICC values of the control group ($M = .55$, $SD = .49$; $t(29) = 2.32$, $p = .027$; $\eta^2 = .16$).

The question arises as to whether similar findings would be seen at other electrodes. To address this, we reanalysed the Time 2 data using the average activity at electrodes Fz, FCz, F3, F4, FC3 and FC4. The relevant data are shown in Figures 3 and 4. The Fisher transformed ICC of the SLI group ($M = .13$, $SD = .37$) was significantly lower than the transformed ICC values of the control group ($M = .58$, $SD = .45$; $t(29) = 3.04$, $p = .005$, $\eta^2 = .242$). It is worth noting that those cases of SLI who had the highest values of ICC tended to be the oldest in their age group. This was not the case for the control group.

One might expect that low ICCs would be found only in cases with poor FD, but, as had previously been observed by McArthur and Bishop (2004), we found several cases in the SLI group who had low ICCs despite good-FD (see Figures 3 and 4).

Do ERPs of Individuals with SLI resemble ERPs of Younger Controls?

Our third aim was to see whether the older cases with SLI had waveforms that were immature for their age. To test this, for those aged 14.5 years or more, ICCs were computed between the individual's waveform and that of the younger control group. (This kind of analysis was not feasible with the younger SLI group, because we did not have even younger controls with whom to compare them). The average waveform from the six frontal electrodes was used. Seven of the ten SLI cases and three of the nine controls had an ICC with the younger grand average that was higher than that with the age-appropriate grand average, and hence may be regarded as having immature auditory ERPs. These cases are labelled with an exclamation mark in figure 4. The trend was in the direction predicted by an

immaturity hypothesis (i.e. a higher proportion of SLI cases with immature waveforms compared with controls), but did not reach statistical significance (Fisher exact test, one-sided $p = .128$). Note, however, that two of the three controls who had immature ERPs were aged 14 years, and thus close to the boundary between young and old groups, and so could be regarded as having a relatively mild maturational lag. In contrast, five of the SLI cases with immature ERPs were 15 years or older, and hence several years older than the younger comparison group that they resembled.

What Causes weak ICCs in the SLI Group?

The ICC data confirm that differences between waveforms of participants with SLI and controls are still present 18 months after the initial assessment. What is responsible for these differences? A low ICC could simply be a consequence of noisy data. Alternatively, it may reflect differences in the amplitude of two waveforms, differences in their shape, or in peak latencies. Additional analyses were carried out to explore these possibilities.

Reliability of Auditory ERPs

Additional analyses were carried out to discover whether the failure to find age-appropriate ICCs in the SLI group might be a reflection of noisy ERP data in the SLI group. We considered four sources of evidence: first, the number of epochs contributing to the waveform; second, the signal-to-noise ratio; third, the split-half reliability of waveforms, and fourth, the reliability of the waveform across Time 1 and Time 2.

The first point to establish was whether the SLI group had noisier data because of a high level of artefact rejection. As shown in Table II, there was a significant difference in number of epochs accepted between SLI and control groups. However, when number of epochs was entered as a covariate, the difference between ICC values for SLI and control groups remained significant, $F(1, 28) = 6.13$, $p = .020$, $\eta^2 = .179$, indicating that the difference was not explicable in terms of this factor. Table II also shows mean values for a measure of signal-to-noise ratio, which was computed by taking the variance of the data points during the baseline (-50 to 0 ms), dividing by the variance of data points post-stimulus (0 to 486 ms), and then taking the natural logarithm of this ratio. SLI and control groups did not differ on this index.

Another indicator of reliability is the consistency of the waveform across the test session. To obtain a measure of split-half reliability within the Time 2 test session we computed waveforms for the 600 Hz and 700 Hz standards separately. On visual inspection, the grand averages for these two waveforms were similar for both SLI and control groups. We used the ICC to provide an index of waveform reliability for each individual participant, again using the Fisher z transform prior to analysis. The means, shown in Table II, do not differ significantly.

The third indicator of waveform reliability was test-retest reliability from Time 1 to Time 2. Relevant data were available for 11 SLI cases and 13 controls, for each of whom the waveforms were compared for the whole epoch at Time 1 and Time 2. Both SLI and control cases had reasonably stable ERP waveforms over an 18 month period, and they did not differ in the reliability of waveforms over time as assessed by the ICC (see Table II).

Overall, all four analyses indicate that the low ICCs found in many of the SLI cases are not simply a consequence of noisy data.

Amplitude

A second possibility was that low ICCs in the SLI group might reflect differences from the control group in amplitude of the ERP. To test this possibility, root mean square (RMS) amplitude at Time 2 was computed for each individual for the interval 100 to 228 ms post onset, again using the mean waveform from the six frontal electrodes, as this gave the largest group effect on ICC comparisons. The RMS amplitude for the SLI group ($M = 10.08$, $SD = 5.70$) did not differ significantly from that of the control group ($M = 8.57$, $SD = 4.43$; $t(29) = 0.82$, $p = .420$). RMS amplitude was, however, significantly correlated with age ($r = -.493$, $df = 29$, $p = .005$).

Waveform Shape

To investigate whether differences in waveform shape were responsible for the relatively low ICCs in the SLI group, we repeated the analysis on the 6-electrode average waveforms from Time 2 using Pearson correlations rather than ICCs. Because the Pearson correlation is sensitive to shape but not amplitude, we would expect this comparison to yield similar results to the ICC analysis if differences in waveform shape were principally responsible for the group difference. For each individual, the waveform between 100 and 228 ms post-onset was compared with the grand average waveform of controls of the same age range. The mean Pearson correlation for group SLI was .50 (Fisher z transformed $M = .65$, $SD = .480$) and for the control group was .71 (Fisher z transformed $M = 1.16$, $SD = .704$). This difference was significant on t -test, $t(29) = 2.33$, $p = .027$, $\eta^2 = .157$, although the effect size was weaker than for the ICC comparison. This analysis suggested that differences in waveform shape, regardless of amplitude, contributed to the observed difference in ICC.

Latency of Peaks

A low correlation (either ICC or Pearson r) could reflect either a difference in the pattern of peaks and troughs in the waveform, or could arise if two waveforms of similar shape were misaligned in time, i.e. if the positive and negative peaks occurred earlier or later than in the comparison waveform. To explore the latter possibility, Pearson correlations were recomputed for Time 2 data, adding or subtracting a time offset. For instance, with an offset of -20 ms, the individual's waveform for the time window 80 to 208 ms would be compared with the grand control average for the same age from 100 to 228 ms. This procedure was repeated in 4 ms steps for offsets ranging from -40 to $+40$ ms for each individual, to identify the offset at which the Pearson correlation reached a maximum. The two groups did not differ significantly in terms of the offset at which the correlation was maximal [for SLI, $M = -4.3$ ms, $SD = 19.67$, for control, $M = 0.8$ ms, $SD = 13.87$; $t(29) = 0.82$, $p = .419$], indicating that there was no consistency from one case to the next in terms of the size or direction of offset that was required to achieve this effect. Thus, there was no evidence for general slowing of ERP components in SLI.

Discussion

Our first aim was to determine whether poor FD thresholds observed by McArthur and Bishop (2004) in a subset of children with SLI would persist over 18 months, or whether thresholds would improve as children grew older. Overall, the results were consistent with the cross-sectional data from Wright and Zecker (2004) and Hautus et al. (2003) in showing that auditory deficits improved with age. Two of the SLI cases who had very high thresholds at Time 1 had normalised by Time 2, and the other two showed improvement in their thresholds. If we had only the Time 1 behavioural data, we would be tempted to conclude that children with auditory frequency discrimination difficulties formed a distinct subgroup of the SLI population. However, the longitudinal data from Time 1 and Time 2 suggest that

deficits in auditory processing may correspond to a delay in auditory maturation rather than a permanent deficit.

Our second aim was to test the replicability of the observation by McArthur and Bishop (2004) that a high proportion of cases of SLI had age-inappropriate waveforms in the N1-P2-N2 region. This result was replicated at Time 2, with a significant difference in the mean ICC for SLI versus control groups. One might have anticipated that auditory ERPs would be abnormal only in those cases of SLI who had poor FD. However, at both times, age-inappropriate ERPs characterised a high proportion of SLI cases, regardless of their FD thresholds. Supplementary analyses revealed that the low ICCs were not due to noisier ERPs in the SLI group. In general, differences in waveform shape rather than differences in amplitude were largely responsible for the ICC difference between individuals with SLI and the control grand mean.

An additional point we would stress is that the SLI cases did not appear to be homogeneous in their ERP waveforms. In particular, there was no evidence for a general slowing of ERP components in the SLI group, as might have been predicted from a temporal processing account. As can be seen by inspection of Figures 3 and 4, there was considerable variability in waveforms from one SLI case to the next.

The third aim of the study was to use ERP data to evaluate further the notion that SLI might involve a maturational lag in auditory processing. The logic of our analysis was similar to that widely used in studies of language and cognition, where a group with SLI is compared not only with an age-matched control group, but also with a younger control group matched on some index such as language age (e.g. Rice et al., 1995). The argument is that if SLI corresponds to some kind of developmental delay, then the profile of performance should resemble that of a younger typically-developing child. In contrast, if the profile does not resemble that of a normal child at any stage of development, it would be more appropriate to treat SLI as a case of deviant development. In the current study, we considered whether older individuals with SLI resembled younger typically developing individuals in their neurophysiological responses to sounds. Our study was small and the data are far from conclusive, but they suggest that a maturational hypothesis may apply to at least a subset of cases of SLI.

Future Questions

When taken in conjunction with behavioural evidence for improvement of auditory functions in SLI, both in the current study, and in research reviewed in the introduction, these data suggest a number of questions for future research. First, for those people with SLI whose ERPs appear immature, we can ask whether this similarity to younger children applies to characteristics of the ERP other than waveform shape and amplitude. For instance, we know that there are age effects on the spatial distribution of activity across electrodes (Bruneau et al., 1997), and on the extent to which the N1 component responds to presentation rate (Ceponiene et al., 2002.). If the brain of the child with SLI is simply immature, we should expect them to resemble younger children in these respects too.

Second, it will of interest to explore individual differences more thoroughly. Three of the cases in the older SLI group had waveforms that were not typical either for their own age group, or for younger children. It is possible that these are cases of more severe delay who would have appeared immature had we had an even younger control group against which to compare them. However, it is also possible that they are an etiologically distinct group. It would be of considerable interest to see whether these cases are more likely than others to show the kinds of minor abnormalities of brain structure described by Leonard et al. (2002) in SLI.

Third, we may consider the relationship between ERP indicators of brain immaturity and other evidence of maturational lag. Although children with SLI, by definition, show normal cognitive development on tests of nonverbal intelligence, they do frequently have deficits on tests of motor function. The link between motor and language impairments appears to reflect common genetic influence on brain regions that are involved in motor and speech processing (Bishop, 2002; see also Lai et al., 2003). On the basis of longitudinal data on motor development, Bishop and Edmundson (1987) argued for a maturational lag account of SLI. As with the auditory deficits described here, they found that motor impairment in young children with SLI improved as they grew older. If we find that auditory and motor indices of immaturity are correlated, this suggests brain regions that mediate both sensory and motor aspects of development are implicated in SLI.

Methodological Implications

Even if a maturational account proves to be invalid, the high rate of age-inappropriate waveforms seen in our SLI cases is of considerable interest. These abnormalities of the ERP would almost certainly have been missed by more traditional methods of analysis that rely on comparison of group means. We focused on studying waveforms at the level of the individual, rather than the group grand mean, because of our concern about the heterogeneity of SLI, using the ICC to compare an individual's waveform with the grand average of a control group. A global measure such as ICC will only tell us how closely two waveforms resemble one another, and will not pinpoint the source of any differences. Nevertheless, just as an overall F-test is done first in an analysis of variance before doing more specific comparisons between conditions, the ICC provides a useful first step when comparing an individual with a control group. It overcomes the difficulties of multiple statistical comparisons that arise when comparing amplitudes and latencies of several different components, and also makes it possible to include data from children who may not show some of the components of interest.

Our study also illustrates how ERP data can offer a source of evidence to complement behavioural data when evaluating a maturational hypothesis of SLI. It emphasises the importance of taking age into account when comparing cases of SLI with control groups on either electrophysiological or behavioural measures that show maturational trends. It suggests that a method widely used in cognitive/linguistic studies, that of comparing SLI cases not just with age-matched controls but also with younger controls who are matched on a measure of language or literacy skill, can help illuminate the underlying nature of any deficits observed.

However, to adopt this approach adequately, one would need a large normative database of waveforms from children of different ages. Gathering such data in a standard ERP paradigm so that it could be shared across different laboratories should be a high priority for the research community. Developmental language and literacy impairments are common but heterogeneous conditions, and for ERP research to throw light on these conditions, we need to progress to the point where we can evaluate an individual's waveform in relation to a age-matched control group in the same way that we currently evaluate language or reading ability in relation to a standardized test.

Acknowledgments

This research was funded by the Wellcome Trust. We thank all our volunteers and their parents for their selfless contributions to this research. We also like to thank Mrs Faith Ayre for co-ordinating our volunteers, and Mr Paul Sutcliffe for help with data collection.

References

- ADAMS J, COURCHESNE E, ELMASIAN RO and LINCOLN AJ. Increased amplitude of the auditory P2 and P3b components in adolescents with developmental dysphasia. In R Johnson, JW Rohrbaugh, and R Parasuraman (Eds), *Current Trends in Event-related Potential Research (EEG and Clinical Neurophysiology*, Suppl 40). New York: Elsevier, 1987, pp. 577–583.
- Ahissar M, Protopapas A, Reid M, Merzenich MM. Auditory processing parallels reading abilities in adults. *Proceedings of the National Academy of Sciences*. 2000; 97:6832–6837.
- Albrecht R, von Suchodoletz W, Uwer R. The development of auditory evoked dipole source activity from childhood to adulthood. *Clinical Neurophysiology*. 2000; 111:2268–2276. [PubMed: 11090781]
- Amitay S, Ahissar M, Nelken I. Auditory processing deficits in reading disabled adults. *Journal of the Association for Research in Otolaryngology*. 2002; 3:302–320. [PubMed: 12382105]
- Baldeweg T, Richardson A, Watkins S, Foale C, Gruzeliier J. Impaired auditory frequency discrimination in dyslexia detected with mismatch evoked potentials. *Annals of Neurology*. 1999; 45:495–503. [PubMed: 10211474]
- Bernstein LE, Stark RE. Speech perception development in language-impaired children: A 4-year follow-up study. *Journal of Speech and Hearing Disorders*. 1985; 50:21–30. [PubMed: 3974208]
- Bishop DVM. Motor immaturity and specific speech and language impairment: Evidence for a common genetic basis. *American Journal of Medical Genetics: Neuropsychiatric Genetics*. 2002; 114:56–63. [PubMed: 11840507]
- Bishop DVM, Bishop SJ, Bright P, James C, Delaney T, Tallal P. Different origin of auditory and phonological processing problems in children with language impairment: evidence from a twin study. *Journal of Speech, Language and Hearing Research*. 1999a; 42:155–168.
- Bishop DVM, Carlyon RP, Deeks JM, Bishop SJ. Auditory temporal processing impairment: Neither necessary nor sufficient for causing language impairment in children. *Journal of Speech, Language and Hearing Research*. 1999b; 42:1295–1310.
- Bishop DVM, Edmundson A. Specific language impairment as a maturational lag: Evidence from longitudinal data on language and motor development. *Developmental Medicine and Child Neurology*. 1987; 29:442–459. [PubMed: 2445609]
- Breier JJ, Gray LC, Fletcher JM, Foorman B, Klaas P. Perception of speech and nonspeech stimuli by children with and without reading disability and attention deficit hyperactivity disorder. *Journal of Experimental Child Psychology*. 2002; 82:226–250. [PubMed: 12093108]
- BRUNEAU N and GOMOT M. Auditory evoked potentials (N1 wave) as indices of cortical development. In B Garreau (Ed), *Neuroimaging in Child Neuropsychiatric Disorders* Paris: Springer-Verlag, 1998, pp. 113–123.
- Bruneau N, Roux S, Guerin P, Barthelemy C. Temporal prominence of auditory evoked potentials (N1 wave) in 4–8-year-old children. *Psychophysiology*. 1997; 34:32–38. [PubMed: 9009806]
- Byring R, Järvillehto T. Auditory and visual evoked potentials of schoolboys with spelling disabilities. *Developmental Medicine and Child Neurology*. 1985; 27:141–148. [PubMed: 3996771]
- Cacace AT, McFarland DJ, Ouimet JR, Schrieber EJ, Marro P. Temporal processing deficits in remediation-resistant reading-impaired children. *Audiology and Neuro Otology*. 2000; 5:83–97. [PubMed: 10720824]
- Ceponiene R, Cheour M, Näätänen R. Interstimulus interval and auditory event-related potentials in children: Evidence for multiple generators. *Electroencephalography and Clinical Neurophysiology: Evoked Potentials*. 1998; 108:345–354. [PubMed: 9714376]
- Ceponiene R, Rinne T, Näätänen R. Maturation of cortical sound processing as indexed by event-related potentials. *Clinical Neurophysiology*. 2002; 113:870–882. [PubMed: 12048046]
- DARUNA JH and RAU AE. Development of the late components of auditory brain potentials from early childhood to adulthood. In R Johnson, JW Rohrbaugh and R Parasuraman (Eds), *Current Trends in Event-related Potential Research (EEG and Clinical Neurophysiology, Suppl 40)* New York: Elsevier, 1987, pp. 590–595.
- DUNN LM, DUNN LM, WHETTON C and PINTILIE D. *British Picture Vocabulary Scales*. Windsor: NFER-Nelson Publishing Co., 1982.

- Escera C, Grau C. Short-term replicability of the mismatch negativity. *Electroencephalography and Clinical Neurophysiology: Evoked Potentials*. 1996; 100:549–554. [PubMed: 8980419]
- Habib M. The neurobiological basis of developmental dyslexia: An overview and working hypothesis. *Brain*. 2000; 123:2373–2399. [PubMed: 11099442]
- Hall JW, Grose JH. The masking-level difference in children. *Journal of the American Academy of Audiology*. 1990; 1:81–88. [PubMed: 2132590]
- Hari R, Saaskilahti A, Helenius P, Uutela K. Non-impaired auditory phase locking in dyslexic adults. *Neuroreport*. 1999; 10:2347–2348. [PubMed: 10439461]
- Hartley DEH, Wright BA, Hogan SC, Moore DR. Age-related improvements in auditory backward and simultaneous masking in 6- to 10-year-old children. *Journal of Speech, Language, and Hearing Research*. 2000; 43:1402–1415.
- Hautus MJ, Setchell GJ, Waldie KE, Kirk IJ. Age-related improvements in auditory temporal resolution in reading-impaired children. *Dyslexia*. 2003; 9:37–45. [PubMed: 12625375]
- Heath SM, Hogben JH, Clark CD. Auditory temporal processing in disabled readers with and without oral language delay. *Journal of Child Psychology and Psychiatry*. 1999; 40:637–647. [PubMed: 10357169]
- Helenius P, Salmelin R, Richardson U, Leinonen S, Lyytinen H. Abnormal auditory cortical activation in dyslexia 100 msec after speech onset. *Journal of Cognitive Neuroscience*. 2002; 14:603–617. [PubMed: 12126501]
- Hill NI, Bailey PJ, Griffiths YM, Snowling MJ. Frequency acuity and binaural masking release in dyslexic participants. *Journal of the Acoustical Society of America*. 1999; 106:L53–L58. [PubMed: 10615710]
- Hogan SC, Moore DR. Impaired binaural hearing in children produced by a threshold level of middle ear disease. *Journal of the Association for Research in Otolaryngology*. 2003; 4:23–29.
- Holopainen IE, Korpilahti P, Juottonen K, Lang H, Sillanpaa M. Attenuated auditory event-related potential (mismatch negativity) in children with developmental dysphasia. *Neuropediatrics*. 1997; 28:253–256. [PubMed: 9413003]
- Holopainen IE, Korpilahti P, Juottonen K, Lang H, Sillanpaa M. Abnormal frequency mismatch negativity in mentally retarded children and in children with developmental dysphasia. *Journal of Child Neurology*. 1998; 13:178–183. [PubMed: 9568762]
- Karhu J, Herrgård E, Pääkkönen A, Luoma L, Airaksinen E, Partanen J. Dual cerebral processing of elementary auditory input in children. *Neuroreport*. 1997; 8:1327–1330. [PubMed: 9172129]
- Kathmann N, Frodl-Bauch T, Hegerl U. Stability of the mismatch negativity under different stimulus and attention conditions. *Clinical Neurophysiology*. 1999; 110:317–323. [PubMed: 10210621]
- Korpilahti P. Auditory discrimination and memory functions in SLI children: A comprehensive study with neurophysiological and behavioural methods. *Scandinavian Journal of Logopedics and Phoniatrics*. 1995; 20:131–139.
- Korpilahti P, Lang HA. Auditory ERP components and mismatch negativity in dysphasic children. *EEG and Clinical Neurophysiology*. 1994; 91:256–264.
- Lai CSL, Gerrelli D, Monaco AP, Fisher SE, Copp AJ. FOXP2 expression during brain development coincides with adult sites of pathology in a severe speech and language disorder. *Brain*. 2003; 126:2455–2462. [PubMed: 12876151]
- Leonard CM, Lombardino LJ, Walsh K, Eckert MA, Mockler JL, Rowe LA, Williams S, DeBose CB. Anatomical risk factors that distinguish dyslexia from SLI predict reading skill in normal children. *Journal of Communication Disorders*. 2002; 35:501–531. [PubMed: 12443050]
- Leppänen PH, Lyytinen H. Auditory event-related potentials in the study of developmental language-related disorders. *Audiology and Neurotology*. 1997; 2:308–340.
- Lincoln AJ, Courchesne E, Harms L, Allen M. Sensory modulation of auditory stimuli in children with autism and receptive developmental language disorder – event related potential evidence. *Journal of Autism and Developmental Disorders*. 1995; 25:521–539. [PubMed: 8567597]
- Lincoln AJ, Dickstein P, Courchesne E, Elmasian R, Tallal P. Auditory processing abilities in non-retarded adolescents and young adults with developmental receptive language disorder and autism. *Brain and Language*. 1992; 43:613–622. [PubMed: 1483193]

- Marler JA, Champlin CA, Gillam RB. Auditory memory for backward masking signals in children with language impairment. *Psychophysiology*. 2002; 39:767–780. [PubMed: 12462505]
- Marshall CM, Snowling M, Bailey PJ. Rapid auditory processing and phonological ability in normal readers and readers with dyslexia. *Journal of Speech, Language and Hearing Research*. 2001; 44:925–940.
- Mason SM, Mellor DH. Brain-stem middle latency and late cortical evoked potentials in children with speech and language disorders. *Electroencephalography and Clinical Neurophysiology*. 1984; 59:297–309. [PubMed: 6203719]
- McAnally K, Stein JF. Auditory temporal coding in dyslexia. *Proceedings of the Royal Society of London, Series B*. 1996; 263:961–965. [PubMed: 8805833]
- McArthur GM, Bishop DVM. Auditory perceptual processing in people with reading and oral language impairments: Current issues and recommendations. *Dyslexia*. 2001; 7:150–170. [PubMed: 11765982]
- McArthur GM, Bishop DVM. Which people with specific language impairment have auditory processing deficits? *Cognitive Neuropsychology*. 2004; 21:79–94. [PubMed: 21038192]
- McArthur G, Bishop DVM, Proudfoot M. Do video sounds interfere with auditory event-related potentials? *Behavior Research Methods, Instruments and Computers*. 2003; 35:329–333.
- Moore JK. Maturation of human auditory cortex: Implications for speech perception. *Annals of Otolaryngology, Rhinology and Laryngology*. 2002; 111(part 2 Suppl):7–10.
- Näätänen R, Alho K. Mismatch negativity – a unique measure of sensory processing in audition. *International Journal of Neuroscience*. 1995; 80:317–337. [PubMed: 7775056]
- Näätänen R, Winkler I. The concept of auditory stimulus representation in cognitive neuroscience. *Psychological Bulletin*. 1999; 125:826–859. [PubMed: 10589304]
- Neville HJ, Coffey SA, Holcomb PJ, Tallal P. The neurobiology of sensory and language processing in language-impaired children. *Journal of Cognitive Neuroscience*. 1993; 5:235–253.
- O'Donnell BF, Shenton ME, McCarley RW, Faux SF, Smith RS, Salisbury DF, Nestor PG, Pollack SD, Kikinis R, Jolesz FA. The auditory N2 component in schizophrenia: Relationship to MRI temporal lobe gray matter and to other ERP abnormalities. *Biological Psychiatry*. 1993; 34:26–40. [PubMed: 8373937]
- Ors M, Lindgren M, Blennow G, Nettelbladt U, Sahlen B, Rosen I. Auditory event-related potentials in children with specific language impairment. *European Journal of Paediatric Neurology*. 2002; 6:47–62. [PubMed: 11993955]
- Pang EW, Taylor MJ. Tracking the development of the N1 from age 3 to adulthood: An examination of speech and nonspeech stimuli. *Clinical Neurophysiology*. 2000; 111:388–397. [PubMed: 10699397]
- Pekkonen E, Rinne T, Näätänen R. Variability and replicability of the mismatch negativity. *Electroencephalography and Clinical Neurophysiology: Evoked Potentials*. 1995; 96:546–554. [PubMed: 7489676]
- Pinkerton E, Watson DR, McClelland RJ. A neurophysiological study of children with reading, writing and spelling difficulties. *Developmental Medicine and Child Neurology*. 1989; 31:569–581. [PubMed: 2806738]
- Ponton CW, Eggermont JJ, Kwong B, Don M. Maturation of human central auditory system activity: Evidence from multi-channel evoked potentials. *Clinical Neurophysiology*. 2000; 111:220–236. [PubMed: 10680557]
- RAVEN JC, COURT JH and RAVEN J. *Raven's Progressive Matrices and Raven's Coloured Matrices*. London: HK Lewis, 1986.
- Reed MA. Speech perception and the discrimination of brief auditory cues in reading disabled children. *Journal of Experimental Child Psychology*. 1989; 48:270–292. [PubMed: 2794857]
- Renvall H, Hari R. Auditory cortical responses to speech-like stimuli in dyslexic adults. *Journal of Cognitive Neuroscience*. 2002; 14:757–768. [PubMed: 12167260]
- Rice ML, Wexler K, Cleave PL. Specific language impairment as a period of extended optional infinitive. *Journal of Speech and Hearing Research*. 1995; 38:850–863. [PubMed: 7474978]
- Rosen S. Auditory processing in dyslexia and specific language impairment. Is there a deficit? What is its nature? Does it explain anything? *Journal of Phonetics*. 2003; 31:509–527.

- Scherg M, Von Cramon D. Evoked dipole source potentials of the human auditory cortex. *Electroencephalography and Clinical Neurophysiology*. 1986; 65:344–360. [PubMed: 2427326]
- Schulte-Körne G, Deimel W, Bartling J, Remschmidt H. Auditory processing and dyslexia: Evidence for a specific speech processing deficit. *Neuroreport*. 1998; 9:337–340. [PubMed: 9507979]
- SEMEL EM, WIG EH and SECORD W. *Clinical Evaluation of Language Fundamentals – Revised* San Antonio: Psychological Corporation, 1987.
- Sharma A, Kraus N, McGee TJ, Nicol TG. Developmental changes in P1 and N1 central auditory responses elicited by consonant-vowel syllables. *Electroencephalography and Clinical Neurophysiology: Evoked Potentials*. 1997; 104:540–545. [PubMed: 9402896]
- Shelton BR, Scarrow I. Two-alternative versus three-alternative procedures for threshold estimation. *Perception and Psychophysics*. 1984; 35:285–392.
- TALLAL P and PIERCY M. Defects of auditory perception in children with developmental dysphasia. In M Wyke (Ed), *Developmental Dysphasia* London: Academic Press, 1978, pp 63–84.
- Tallal P, Stark R, Kallman C, Mellits D. A re-examination of some nonverbal perceptual abilities of language-impaired and normal children as a function of age and sensory modality. *Journal of Speech and Hearing Research*. 1981; 24:351–357. [PubMed: 7300276]
- Taylor MJ, Baldeweg T. Application of EEG, ERP and intracranial recordings to the investigation of cognitive functions in children. *Developmental Science*. 2002; 5:318–334.
- Thompson NC, Cranford JL, Hoyer E. Brief-tone frequency discrimination by children. *Journal of Speech, Language, and Hearing Research*. 1999; 42:1061–1068.
- Tonnquist-Uhlén I. Topography of auditory evoked cortical potentials in children with severe language impairment: The P2 and N2 components. *Ear and Hearing*. 1996a; 17:314–326.
- TONNQUIST-UHLÉN I. Topography of auditory evoked cortical potentials in children with severe language impairment. *Scandinavian Audiology*, Supplement 44, 1996.
- Uwer R, Albrecht R, von Suchodoletz W. Automatic processing of tones and speech stimuli in children with specific language impairment. *Developmental Medicine and Child Neurology*. 2002; 44:527–532. [PubMed: 12206618]
- Uwer R, Von Suchodoletz W. Stability of mismatch negativities in children. *Clinical Neurophysiology*. 2000; 111:45–52. [PubMed: 10656510]
- Waber DP, Weiler MD, Wolff PH, Bellinger D, Marcus DJ, Ariel R, Forbes P, Wypij D. Processing of rapid auditory stimuli in school-age children referred for evaluation of learning disorders. *Child Development*. 2001; 72:37–49. [PubMed: 11280488]
- WIG EH and SECORD W. *Test of language competence: expanded edition* San Antonio: Psychological Corporation, 1989.
- Wright BA, Zecker SG. Learning problems, delayed development, and puberty. *Proceedings of the National Academy of Sciences*. 2004; 101:9942–9946.
- Wright BA, Bowen RW, Zecker SG. Nonlinguistic perceptual deficits associated with reading and language disorders. *Current Opinion in Neurobiology*. 2000; 10:482–486. [PubMed: 10981617]
- Wright BA, Lombardino LJ, King WM, Puranik CS, Leonard CM, Merzenich MM. Deficits in auditory temporal and spectral resolution in language-impaired children. *Nature*. 1997; 387:176–178. [PubMed: 9144287]

Fig. 1.
The auditory ERP of the oldest control participant tested at Time 2, electrode Fz.

Fig. 2.

Scatterplot showing relationship between FD thresholds of participants in the SLI group and control group at Time 1 and Time 2. The threshold corresponds to the frequency at which the higher tone had to be set to be distinguished from a 600 Hz standard tone. The four participants with thresholds over 700 Hz at time 1 constituted the original poor-FD group. Points falling on the dotted line correspond to cases obtaining the same score at Time 1 and Time 2. Those falling below the line show improvement from Time 1 to Time 2.

Fig. 3.

Individual auditory ERPs (average of electrodes Fz, FCz, F3, F4, FC3 and FC4) for younger participants with SLI (left column) and controls (right column) with dotted grey line indicating age-appropriate grand average comparison waveform for controls (see text). Significance of ICCs (one-tailed) denoted by * = .05, ** = .025 and *** = .01. Shading denotes cases from original SLI-poor FD group.

Fig. 4.

Individual auditory ERPs (average of electrodes Fz, FCz, F3, F4, FC3 and FC4) for older participants with SLI (left column) and controls (right column) with dotted grey line indicating age-appropriate grand average comparison waveform for controls (see text). Significance of ICCs (one-tailed) denoted by * = .05, ** = .025 and *** = .01. Shading denotes cases from original SLI-poor FD group. ! denote case where ICC with younger controls is higher.

Mean Non-verbal IQ, and Language Test Scaled Scores of SLI and Control Individuals who were Retested at Time 2

TABLE 1

	SLI (N = 16)			Control (N = 15)		
	Mean	SD	Range	Mean	SD	Range
Non-verbal ability ¹	91.87	13.36	75–113	98.13	9.69	75–113
BPVS ²	79.12	15.88	56–110	110.00	10.02	93–128
Figurative Language ³	4.38	1.78	3–9	10.80	2.86	7–16
Recreating Sentences ³	4.31	2.15	3–6	8.33	2.16	6–13
Recalling Sentences ⁴	4.25	1.34	3–10	10.27	2.05	6–13

¹ Standard Progressive Matrices (Raven, Raven, and Court, 1998), scaled with a mean of 100 and SD of 15.

² British Picture Vocabulary Scale (Dunn et al., 1982), scaled with a mean of 100 and SD of 15.

³ Subtest from Test of Language Competence-Expanded Edition (TLC-EE; Wiig and Secord, 1989), scaled with a mean of 10 and SD of 3.

⁴ Subtest from Clinical Evaluation of Language Fundamentals-Revised (Semel et al., 1986), scaled with a mean of 10 and SD of 3.

TABLE II
Mean (SD) Reliability Indices for ERP Waveforms in SLI and Control Groups at Time 2

	SLI (N = 16)		Control (N = 15)		t (29)	p
	M	SD	M	SD		
Epochs accepted	1310.3	91.18	1371.0	36.37	2.41	.023
SNR ^a	4.46	1.63	4.62	1.03	0.31	.762
Split-half reliability ^b	.69	.158	.77	.190	1.19	.243
Test-retest reliability ^{c,d}	.81	.380	.77	.580	0.18	.860

^aLn(variance prestimulus/variance poststimulus)

^bFisher transformed ICC between 600 Hz vs 700 Hz standards (whole waveform).

^cMeans are based on N = 11 for SLI and N = 13 for controls.

^dFisher transformed ICC between Time 1 and Time 2 (whole waveform).