**Test-retest reliability of Mismatch Negativity and Late Discriminative Negativity response in children with listening difficulties**

Short title: MMN and LDN in children

Mridula Sharma1,2, Varghese Peter3, Danielle Dennis 2, Gitanjali Raman4, Suzanne C Purdy5,6

1 College of Nursing and Health Sciences, Flinders University, Adelaide, Australia

2 Department of Linguistics,Macquarie University, Sydney, Australia

3 School of Health, University of the Sunshine Coast, Queensland, Australia

4 Genetics, Cell Biology and Development, University of Minnesota, Twin Cities, United States

5 School of Psychology, University of Auckland, New Zealand

6 Eisdell Moore Centre for Hearing and Balance Research, University of Auckland, New Zealand

**Corresponding author:** Mridula Sharma, PhD

College of Nursing and Health Sciences  
Flinders University, SA, Australia  
Email: [mridula.sharma@flinders.edu.au](mailto:mridula.sharma@flinders.edu.au)

**Paper word count:** 6803

**Figure count:** 2

**Table count:** ~~3~~ 4

**Supplementary Table**: 1

**Abstract**

Mismatch Negativity (MMN) and Late Discriminative Negativity (LDN) event related potentials are objective indicators of auditory discrimination. The aim of the study was to determine the test-retest reliability of MMN and LDN recorded to simple speech contrasts in children with listening difficulties**.** MMN and LDN responses were recorded from Fz and Cz electrodes for a /da/-/ga/ contrast twice within a 10-day period. Intraclass Correlation coefficients (ICC) were used to determine test-retest reliability of MMN and LDN. Eight five children (55 males, aged 7.0-12.8 years) with listening difficulties participated in this research. Children were grouped into four clusters based on their reading, language, nonverbal intelligence, and cognitive skills such that children within each cluster had similar profiles of strengths and difficulties. ICC for MMN were better than LDN. Results showed that MMN amplitude did not differ between visits, but LDN amplitude reduced significantly (more positive peak amplitude) at visit 2 compared to visit 1. At visit 1, MMN and LDN were detectable in only 41%/42% and 18%/21% (Fz/Cz) of participants respectively. MMN was most replicable (71%) for children with listening difficulties in the cluster with relatively good nonverbal intelligence and language skills. The results do not support the clinical utility of LDN for objective assessment of auditory discrimination. Although MMN had better test-retest reliability, overall detectability was poor; improved detectability is needed for MMN to have robust clinical utility in children.

**Keywords**

Mismatch Negativity; Late Discrimination Negativity; children; listening difficulties

**Abbreviations**

Mismatch Negativity: MMN

Late Discrimination Negativity**:** LDN

Auditory Processing Disorder: APD

**1 INTRODUCTION**

Mismatch Negativity (MMN) and the Late Discriminative Negativity (LDN) are discriminative auditory evoked responses that are not influenced by non-auditory factors such as active attention or motivation, that reflect discrimination of differences between auditory standard and deviant stimuli presented in an oddball sequence (Bishop et al., 2011). MMN and LDN are elicited by irregularities in a stream of repetitive auditory stimulation and can be recorded from infancy (Cheour et al., 2001), making them good candidates for objective measures of auditory discrimination. MMN and LDN are derived by subtracting the auditory evoked response to repetitive standard sounds from the evoked response to infrequently presented deviant sounds (Ruhnau et al., 2012). Both responses are typically measured using paradigms where the participant is not attending to the stimuli and is involved in another activity such as watching a silent movie or reading a book. MMN and LDN are thought to reflect the automatic and pre-attentive detection of stimulus change (Näätänen et al., 2004; Picton et al., 2000). MMN is a consequence of discrimination in some acoustic domain detected within auditory cortex when an infrequently presented stimulus (deviant) contrasts with the memory trace of the frequently presented stimulus (standard) (Fitzgerald and Todd, 2020; Näätänen et al., 2005).

The LDN response has been associated with language-specific (Cheour et al., 2001), word meaning (Korpilahti et al., 2001) or attention-related (Escera et al., 2000a) processes. Like MMN, LDN can be elicited by speech stimuli and complex tones (Zachau et al., 2005), as well as simple auditory contrasts (Hakvoort et al. (2015). The functional significance of LDN is under debate and may depend on the selection of stimuli. The contribution of lexical processing (Cheour et al., 2001; Korpilahti et al., 2001) versus reorientation of attention (after attention is involuntarily given to the deviant stimulus) (Lu et al., 2015; Wetzel and Schröger, 2014) is likely to depend on the choice of stimuli. The LDN is reported to be particularly prominent in children compared to adults (Cheour et al., 2001)*.* Cheour et al.noted bigger LDN amplitudes in 4- and 8-year-olds compared to adults to tonal stimuli indicating a significant effect of age, unlike MMN which is developmentally more stable*.*

MMN is a negative response optimally elicited with frontocentral and central scalp electrode placement (Picton et al., 2000). The LDN is a longer-latency negative response occurring at about 400 ms after stimulus onset with a more central distribution (Korpilahti et al., 1995). Unlike MMN which increases in amplitude with increasing standard-deviant stimulus difference, LDN appears to be larger (i.e., more negative) for smaller stimulus differences (Bishop et al., 2011; Liu et al., 2014).

Both MMN and LDN appear to be sensitive to auditory training effects, however, while MMN amplitudes are shown to improve with training, LDN amplitude changes are not yet clear. Kujala and Leminen (2017) reported a randomised control trial in which 5-year-old children with language impairment were trained on speech discrimination and noted that, post-training, MMN increased in amplitude consistent with the improvement in speech discrimination after training. Pihko et al. (2007) also reported post-training MMN increases, measured using magnetoencephalography. They trained speech discrimination in 6-7-year-old children with language impairment and found that, post-training, MMN increased in amplitude for one of the contrasts /sy/. There were also performance improvements on the behavioural discrimination task for other syllables that were hard to discriminate pre-training. In contrast, in a lexical tone training study, LDN decreased in amplitude post training, possibly indicating that training led to a more efficient attentional reorientation to lexical tone changes (Kaan et al., 2008). Differential effects of training on MMN and LDN have been reported (Putkinen et al., 2014). For example, 4–6-year-old children were given French or music training for four weeks; post training, both training groups showed enhanced LDN response but no change in MMN possibly due to focus on cognitive training rather than specific auditory training (Moreno et al., 2015).

The detectability and stability of the response over time in the absence of auditory training are important considerations when examining training studies in which changes in MMN and LDN are interpreted as evidence for improved discrimination. In studies of neuro-typical people MMN has been identified as “reliable”, however, test-retest correlations vary widely, ranging between 0.37 to 0.87 across studies (Dalebout and Fox, 2001; Escera et al., 2000b; Frodl-Bauch et al., 1997; Kathmann et al., 1999; Pekkonen et al., 1995; Tervaniemi et al., 1999; Wang et al., 2012). The MMN literature often reports high reliability at a group level, but not at an individual level, for both neurotypical and clinical populations (Leppänen et al., 2019). To our knowledge the test-retest reliability of LDN has not been reported.

MMN, more than LDN, has been assessed in clinical populations as well as in typical populations (Bishop et al., 2011). The stimuli used to elicit MMN (and LDN) have ranged from simple tones to complex stimuli including tones and speech tokens across studies. In typical children, previous research on /da-ga/ stimuli have reported presence of MMN in Sharma et al (2004) where 88% of typical children showed MMN to /da-ga/ contrast (*n*=8); using the same paradigm and more participants (*n*=21), 57% of typical children showed MMN to /da-ga/ contrast. Thus, MMN detectability at an individual level has been found to be highly varied and affected by the specific stimulus contrast.

A few studies have reported on test-retest reliability for MMN in adults (Cocquyt et al., 2023) and in children to speech stimuli. In a review of literature of MMN in adults, Cocquyt et al. (2023) highlighted the general reliability for MMN latency but not MMN amplitudes. Uwer and Suchodoletz (2000) investigated the stability of MMN in 15 typically developing children aged 7-11 years to speech stimuli (/da/, /ga/, /ba/) at 350-560 ms. MMN response was only consistently present across conditions in six children (43%); *n*=13 (87%) children showed a present MMN to /ga/ in at least one session.

MMN may be smaller in amplitude/area, later or less detectable in clinical populations compared to controls. For instance, researchers have reported absent (Davids et al., 2011; Kujala and Leminen, 2017) or attenuated (Kujala and Leminen, 2017; Uwer et al., 2002) MMN in children with language impairment. Children with reading disorders show attenuated MMN to speech stimuli (Bishop, 2007). Hakvoort et al. (2015) investigated MMN and LDN in fluent and poor readers and found that there were no MMN differences across groups, but the LDN to intensity differences was only present in controls. Findings are less clear for children with auditory processing disorder (APD) (Hakvoort et al., 2015). A few studies have shown typical MMN in children with APD (Koravand et al., 2017; Liasis et al., 2003; Roggia and Colares, 2008), whereas others have shown reduced MMN in APD (Sharma et al., 2006). Such differences in findings for children with APD may be due to the heterogeneity prevalent within the APD population (Sharma et al., 2019, 2014; Tomlin et al., 2015). It is, therefore, essential to evaluate MMN at an individual level or at least within clusters of children with similar profiles of difficulties (Sharma et al., 2019).

In the current study MMN and LDN were recorded across two visits in children with listening difficulties that fell into four clusters based on hierarchical cluster analysis, as previously reported (Sharma et al., 2019). The four clusters consisted of children that in general had the four patterns of difficulties or strengths: 1) global difficulties on auditory processing, reading, language and cognitive skills, 2) poor auditory processing with good word reading skills, 3) poor auditory processing, attention, and memory but good non-verbal intelligence and language skills, and 4) poor auditory processing and attention with good memory skills. The current research aimed to evaluate the test-retest reliability of MMN and LDN for these four clusters of children on a discriminative evoked potential task that has been used previously to evaluate auditory discrimination in typically developing children (Sharma et al., 2004) and in children with reading disorders (Sharma et al., 2006). MMN was detectable in 57% of control group children with no listening difficulties for the /da/-/ga/ contrast in these studies, which is consistent with other studies that have found similar MMN detection rates (60%, Uwer and Suchodoletz 2000). The detection rates for typical children have been previously reported to be lower than in adults, which could be due to the nature of background EEG activity resulting in poorer evoked response signal to noise ratios in children (Uwer and Suchodoletz 2000).

We hypothesised that MMN and LDN responses recorded under the same conditions within two visits over a 10-day period would be similar in both detectability and response amplitude, but that there would be differences between clusters since different MMN/LDN findings across studies are likely to reflect sample heterogeneity. We hypothesised that Cluster 1 participants would show absent MMN/LDN responses on visit 1, based on earlier evidence for MMN/LDN differences in children with reading, language and/or auditory processing difficulties (Bishop, 2007; Kujala and Leminen, 2017; Sharma et al., 2006).

**2 MATERIALS AND METHODS**

2.1 Participants

Ninety school-aged children aged 7-12 years with listening difficulties participated in the study. Of the 90, eighty-five children (55 males and 30 females) came in for the two visits and are included in the current study. The children aged 7.0 to 12.8 years (mean = 9.78 years ± 1.52 and median 9.8 years) participated in the study (F: 9.71 ±1.54; M: 9.82 ±1.51). All children had hearing within 20 dB HL across all frequencies with normal compliance and peak pressure on the tympanometry on the day of the testing. Based on their performance on auditory, phonological awareness, reading and language tasks, all children were grouped into four clusters as previously presented in Sharma et al. (2019) and provided in Table 1. The study was approved under the University of Auckland Human Research Ethics for basic ethical considerations for the protection of human participants in research.

2.1.1 Clusters

The Test of Nonverbal Intelligence (TONI-3), Queensland Inventory of Literacy (QUIL) subtest of phonemic manipulation, Castles and Coltheart (CC) word/nonword reading task were used to create the four clusters (Sharma et al., 2019). TONI-3 required children to complete a matrix pattern based on shape, direction, position, or size and the test required minimal linguistic input. Phonemic manipulation task required children to determine new word when a phonemic was removed from a word, e.g. “spit” without /p/ is “sit”. The Castles and Coltheart word/nonword test required children to read a list of words (30 were made-up words) and the score of nonword reading were considered here. In addition, CELF–IV was used to measure receptive language and included a digit span (working memory) task. Receptive language is a composite score that included Concepts and Following Directions, Receptive Word Classes, and Sentence Structure subtests. The Digit Span Backward Test required children to listen and repeat numbers in reverse order. For the sustained attention task (IVA Integrated Visual and Auditory Continuous Performance Test), children responded to number 1 while ignoring number 2; this task took about 15 minutes to complete (more details of testing provided in Sharma et al., 2019). To evaluate MMN/LDN test-retest, each participant was tested on the paradigm twice within 10 days.

The cluster analysis, as reported in Sharma et al (2019), had revealed four clusters consisted of children with: 1) global difficulties on auditory processing, reading, language and cognitive skills, *n*= 34; 2) poor auditory processing with good word reading skills, *n*=19; 3) poor auditory processing, attention, and memory but good non-verbal intelligence and language, *n*=14; and 4) poor auditory processing and attention with good memory skills, *n*=18 (Table 1).

INSERT Table 1.

2.2 Stimuli

Short natural speech tokens /da/ and /ga/ (160±4 ms) were used to elicit MMN and LDN (as previously published in Sharma et al., 2004; 2006). The oddball paradigm included /da/ as standard presented 90% with /ga/ as deviant (10%). The speech stimuli, /da/ and /ga/ [/a/ as in the word h**a**rd], were spoken in isolation by an Australian female. The low back ‘ah’ vowel is produced without the ‘r’ sound in the word ‘hard’ in Australian English (Sharma et al., 2004). In addition, /ga/ was also presented 100% of the time in a separate deviant-only (control) stimulus block.

2.3 Paradigm

All children watched a silent movie of their choice with subtitles during the task. They were instructed to not pay attention to the stimuli. The stimuli were presented in an oddball paradigm where the first 10 stimuli were all standard while there were at least 3 standard stimuli in between deviant sounds. There were 2000 standards with 200 deviants, split into three blocks of about 5 minutes each, presented to each participant. Interstimulus stimulus (ISI), the duration from offset of a stimulus to onset of the following stimulus, was 550ms. After the presentation of the three oddball blocks, the deviant stimuli were presented 500 times without the intervening standards (control block).

The continuous EEG was recorded using SynApms2 system (Compumedics Neuroscan). Sounds were presented using STIM software and hardware to an ER-3A insert earphone bilaterally at 70 dB SPL (re: 2cc coupler). EEG was recorded in continuous mode (gain 500, filter 0.1–100 Hz) using SCAN (version 4.2) via gold cup electrodes placed at Fz, and Cz with the reference electrode on the right earlobe and ground on the forehead.

2.4 Offline analysis

The EEG was analysed offline using EEGLAB (Delorme and Makeig, 2004) and ERPLAB (Lopez-Calderon and Luck, 2014) toolboxes running in MATLAB 2019a (Mathworks, Natick, MA, USA). The EEG was first bandpass filtered between 0.1 and 30 Hz (‘pop\_eegfiltnew’ function in EEGLAB). The EEG was then divided into epochs between -100 to 500 ms relative to the sound onset. As there were 500 control stimuli, 200 epochs for the controls were randomly selected to match the number of epochs for deviants. Epochs with amplitude exceeding ± 100 mV were rejected. Prior to averaging, EEG epochs were baseline corrected using the pre-stimulus period (−100 to 0 ms). ~~The average number of accepted epochs were as follows: visit 1 deviants M= 176.33, SD=24.73; visit 1 controls M=182.65; SD=11.73; visit 2 deviants M= 176.59, SD=24.44; visit 2 controls M=181.06; SD=14.33 which amounts to no more than 12% rejection rate.~~ The average numbers of accepted epochs across clusters, stimuli and visits are shown in Table 2.

Insert Table 2

A 3-way mixed ANOVA for the number of accepted epochs with between subject factor Cluster (4 levels: Cluster 1,2,3,4) and within subject factors visit (2 levels: visit 1, visit 2) and stimuli (2 levels: deviants, controls) did not reveal any significant main effects or interactions (all F<3.5, all p<.05), ~~A 2 x 2 repeated measures analysis of variance (ANOVA) on the number of accepted epochs with the factors visit (visit 1, visit 2) and stimuli (deviants, controls) did not show any significant main effects or interactions~~ suggesting no systematic signal-to-noise ratio difference between clusters, stimuli and visits. The epochs were then averaged to derive the ERPs for controls and deviants for each visit. The difference waves were computed by subtracting the ERPs to the control stimuli (deviant alone) from the deviant (within the oddball) responses. The ERPs from individual subjects were averaged to obtain the grand averaged ERPs.

The significance of MMN/LDN in the grand averaged waveform was analysed using a paired t test at each time point after 100 ms for the individual ERP waveforms for deviants and controls. We applied a correction for multiple comparisons using a statistical temporal cluster extent threshold (Guthrie and Buchwald 1991). This approach assesses the autocorrelation across consecutive time points in the ERP signal, allowing us to establish the minimum consecutive time points required to display a statistically significant difference between the deviant and control waveforms using a two-tailed p < .05 significance level, corrected for cluster significance (for further details, see Guthrie and Buchwald 1991).

~~computed by point-by-point t-test between deviant and control ERPs. The difference was considered significant if 12 consecutive datapoints showed a significant t value (Guthrie and Buchwald, 1991).~~ The significance of MMN/LDN was also tested at an individual level using the same method ~~across trials~~ where instead of participants, we used individual trials. This approach for significance testing is common in studies using sparse electrodes (Petit et al., 2020; Vergara-Martínez et al., 2020). The MMN amplitude was calculated from individual subjects in a 50-ms time window around the peak of MMN in the grand averaged waveform. In the absence of a significant peak, the amplitude calculated was used as a measure of noise in the same period. We also calculated the noise levels in the MMN calculation window using the measure aSME (analytic standard measurement error) to qualify the quality of the EEG specifically at latencies where MMN and LDN were calculated, as implemented in ERPLAB (Luck et al., 2021).

The point-by-point t-test between deviants and controls in the group level analysis also showed significant effects in a time range beyond the MMN time window. This response was consistent with the late discriminative negativity (LDN). LDN amplitude was calculated as the mean amplitude within a 350-500 ms time window after the onset of the stimulus as has been previously described (Cheour et al., 2001). aSME was calculated as a measure of noise levels in this range.

**3 RESULTS**

MMN peak latencies at Fz and Cz were noted between 242–292 ms for visit 1 and between 244–294 ms at Fz and 229–279 ms at Cz for visit 2. The aSME values across participants for MMN were 1.76 and 1.73 for visit 1 and 1.73 and 1.69 for visit 2 at the two electrodes (Fz and Cz, respectively). LDN was noted to be a broad negative response and therefore did not necessarily have a peak latency. LDN amplitude and aSME were determined as the mean amplitude within the 350–500ms window at Fz and Cz. The aSME value across participants for LDN was 1.67 at each electrode for visit 1 and 1.71 and 1.70 at the two electrodes (Fz and Cz respectively) for visit 2.

Preliminary analyses using repeated-measures ANOVA with participant age as a covariate (as there are changes in cortical responses between 7-12 years of age) and for visit 1 showed no significant effects of recording electrode (Fz vs Cz) for MMN [F(1, 83)=1.29, *p*=0.26] or LDN [F(1, 83) =1.63, *p*=0.21] (Supplementary Figure 1). To ensure that the results across the two channels are similar, a two-way random effects model of intraclass correlation (ICC) coefficient was determined for MMN and LDN between Cz and Fz for visit 1. ICC has been recommended as a preferred measure of test-retest reliability (Koo and Li, 2016). The ICC coefficients were found to be highly reliable at 0.93 (95% CI 0.90 to 0.96) and 0.93 (95% CI 0.90 to 0.96) respectively. The ICC coefficients were similarly high for visit 2 between Fz and Cz for MMN at 0.93 (95% CI 0.89 to 0.95) and LDN at 0.94 (95% CI 0.91 to 0.96). Given the highly similar results across the two channels, from this point on, only Cz results will be reported.

3.1 Reliability of MMN and LDN across visit 1 and visit 2 for all children (*n*=85)

~~MMN latency at Fz for visit 1 was 267 ms (window of 242-292) and 269 ms (window of 244-294) for visit 2.~~ MMN latency at Cz for visit 1 was 267 ms (window of 242–292) and for visit 2 was 254 ms (window of 229–279 ms). Supplementary Table 1 shows the grand averages of MMN and LDN amplitudes for the two visits. The ICC coefficient between two visits (two-way random effects model) at Cz was found to be moderate for MMN at 0.58 (95% CI 0.35 to 0.73) and poor for LDN at 0.42 (95% CI 0.06 to 0.64). ~~The ICC for Fz showed lower reliability for both MMN at 0.40 (95% CI 0.08 to 0.61) and LDN at 0.36 (95% CI 0.02 to 0.58).~~

Repeated measures ANOVA (2 visits, ~~averaged~~ ~~Fz/Cz electrodes~~) at Cz with age as a covariate showed no visit effect for MMN amplitude [F(1, 83) =0.44 ~~0.17~~, *p*=0.~~68~~51]. A significant visit effect was observed for LDN amplitude [F(1, 83) =10.11 ~~6.41~~, *p*=0.002]; LDN amplitude was significantly more positive, i.e., showed reduced amplitude, for visit 2 compared to visit 1 (Figure 1). There was a between-subject significant age effect ~~only~~for both MMN [F(1,83) = 5.19, *p*=0.03] and LDN [F(1,83) = 12.88, *p*<0.001].

INSERT Figure 1

3.1.1 Differences in MMN and LDN amplitudes across the four clusters

Table ~~2~~3 shows the grand average amplitudes for MMN and LDN at ~~Fz and~~ Cz ~~(and averaged across electrodes)~~ for the four clusters.

INSERT Table ~~2~~3

A 2-way random effects model of ICC coefficients was determined for MMN and LDN at Cz for the two visits for each cluster (Table ~~3~~ 4). The ICC coefficients for both responses were poor but MMN showed relatively better reliability than LDN.

INSERT Table ~~3~~ 4

To explore differences between clusters and across visits, repeated measures ANOVAs were conducted, with age as a co-variate, for MMN and LDN amplitudes (averaged across electrode montage, Table 3). Levene’s tests showed the variance across the groups to be equal for visit 1 for MMN [F(3, 81) = 0.62, *p*=0.61] and LDN [F(3, 81) = 0.15, *p*=0.99] and similarly for visit 2 MMN [F(3, 81) = 0.41, *p*=0.07] and LDN [F(3, 81) = 0.72, *p*=0.54].

Repeated measures ANOVA (2 visits, 4 clusters, age as co-variate) showed no visit [F(3, 80) = 0.01, *p*=0.93] or cluster [F(3, 80) = 1.23, *p*=0.30] effect for MMN. There was a significant overall visit effect for LDN [F(1, 80) = 4.71, *p*=0.03], but no LDN differences across clusters [F(3, 80) = 0.06, *p*=0.98]. There was also no interaction between visit and cluster for LDN amplitude [F(3, 80) = 1.06, *p*=0.37], thus the test-retest differences in LDN were consistent across participants.

3.1.2 Presence of MMN and LDN at the individual level

Table ~~4~~ 5 shows the distribution of the number of participants within each cluster who showed MMN and LDN at ~~Fz and~~ Cz. The difference waves in Figure 2 show MMN and LDN for the two visits across the four clusters. For both visits, more children had measurable MMN compared to LDN (see Supplementary figure for individual waveforms for all clusters for the two visits (red for visit 1 and blue for visit 2). Children in Cluster 3, with relatively good language skills and good nonverbal intelligence scores, showed the most consistent presence of MMN and LDN for both visits (Table 3), and relative to other clusters, inter-individual reliability of MMN and LDN responses based on response detectability was ~~relatively low~~high for Cluster 3.

INSERT Table ~~4~~ 5

INSERT Figure 2

**4 DISCUSSION**

MMN and LDN are discriminatory cortical responses evoked when the deviant sound is perceived as different from the expected, standard stimuli. Previous studies have found that the LDN response is most prominent in response to discrimination of speech sounds (Cheour et al., 2001; Korpilahti et al., 2001), and the /da-ga/ contrast has been used by this study’s authors to elicit MMN in typical children (Sharma et al. 2004). Sharma et al. (2006) found that /ga/ elicited MMN was the smallest in children with reading difficulties compared to other stimuli, hence the /da-ga/ stimulus contrast is a potentially useful speech-contrast for showing individual differences and was used to elicit MMN and LDN in this current study. This contrast has been used in previous studies with moderately reliable presence of MMN at group level (Sharma et al., 2004; Sharma et al., 2006; Bishop, 2007). Consistent with previous findings (Bishop et al., 2011), MMN was observed within 100–300ms and LDN between 350–500 ms after the stimulus for children in all four clusters. Across the four clusters, between 32–71% of participants showed MMN and 15–36% had LDN consistently recorded for both visits (Table 3). The current research fulfills ~~some of~~ identified gap~~s~~ in the MMN literature including ~~reasonable sample size within each cluster and~~ providing individual data. However, one of the limitations of the current study is the lack of test-retest data for a control group of children without listening difficulties. Despite this limitation, the results are consistent with two previous test-retest reliability studies conducted in people with no listening difficulties which also raised concerns for MMN regarding poor across-session reliability in both children (Uwer and Suchodoletz, 2000) and adults (Dalebout and Fox, 2001).

Another limitation of the current study was that only two electrodes Cz and Fz were used. These two electrodes were chosen to reflect the clinical setting and clinical equipment where only two channels are available. Unfortunately, this limited electrode montage does not allow measurement of eyeblink contamination. Previous studies investigating test-retest reliability, like the current research, also used Cz and Fz (Uwer and Suchodoletz, 2000; Dalebout and Fox, 2001). It is possible that perhaps other frontal and/or central sites may be better for capturing MMN and LDN. Future studies may want to consider this aspect when researching or using MMN in clinical populations.

Although at the group level MMN in typical children is reported to be “reliably present”, MMN is highly variable across studies and, to our knowledge, there are no studies that have reported the test-retest reliability of LDN. While the information from typical populations is useful, the reliability of the responses in clinical populations is critical if the responses are to be used to inform diagnosis or management, or to evaluate auditory training effects. In the current research, MMN and LDN were elicited, but not consistently across the two visits. This is an important finding as most clinical studies have not investigated test-retest for MMN in clinical cohorts and hence it is difficult to confidently rely on reported differences in clinical populations, given that the results may vary from session to session. LDN amplitudes reduced significantly for visit 2 while MMN amplitudes, when present, remained similar in size across the two visits, consistent with previous reports of acceptable test-retest reliability for MMN (Escera et al., 2000b; Frodl-Bauch et al., 1997; Kathmann et al., 1999; Pekkonen et al., 1995; Tervaniemi et al., 1999; Wang et al., 2021).

It was hypothesised that MMN and LDN responses recorded under the same conditions within two visits over a 10-day period would be similar in both detectability and response amplitude. This hypothesis was rejected for LDN. There were test-retest differences for LDN amplitude, with smaller LDN at visit 2. The number of individual children with detectable LDN was higher at visit 2 (Table 3), however, the detectability of LDN was generally poor. LDN was present for both visits in only 15% of the children. The LDN differences across the two sessions may be due to learning of the contrasted speech stimuli. The current study used simple speech tokens, and the children watched a silent movie of their choice during EEG recordings. Hence, language specific processing or word meaning were unlikely to contribute to presence of LDN in the current study (Cheour et al., 2001; Escera et al., 2000a; Korpilahti et al., 2001). It may be that the /da-ga/ tokens elicited LDN as this stimulus contrast resulted in involuntary reorienting of the children’s attention to the contrast, as has been previously suggested (Lu et al., 2015; Wetzel and Schröger, 2014).

Differences in LDN between visits cannot be explained by quality of the recordings as the noise levels in the responses measured using aSME for the two groups and two visits did not differ. Attention or motivation also cannot readily account for the differences between visits since the children were engaged in a passive listening task and were consistently engaged in watching a silent self-selected subtitled video during the EEG recordings (Engström et al., 2020; Frey et al., 2019; Leung et al., 2021; Sharma et al., 2006). It is possible that children who showed an LDN response in the current study were showing involuntary reorientation to the speech contrast and, because of repeated testing, they learnt to better ignore the contrast, therefore eliciting a smaller LDN ~~become smaller in visit 2~~. Kaan et al. (2008) reported reduced amplitude of LDN after active training and suggested that smaller LDN amplitudes might be indicative of less reorienting of attention due to the familiarity to the contrast after training. While there was no active training provided in the current study, only passive exposure to the repeated stimuli, it is possible that this also affected the children’s attention to the stimulus contrast, and hence could account for differences in the size and detectability of the LDN at visit 2.

The four clusters were based on distinct areas of strengths and difficulties on auditory and cognitive tasks. MMN was elicited in about half of the children at visit 1. Between 32-71% of participants showed MMN consistently recorded for both visits across the four clusters (Table 3). We hypothesised that MMN detectability and reliability would differ across clusters, based on earlier evidence for MMN differences in children with reading, language and/or auditory processing difficulties (Bishop, 2007; Kujala and Leminen, 2017; Sharma et al., 2006). Poor speech discrimination, a characteristic of APD, is also associated with absent or small MMN in previous literature (Davids et al., 2011; Kujala and Leminen, 2017; Uwer et al., 2002; Kraus et al., 1996). The hypothesised MMN differences across clusters was not supported by our results which showed no MMN amplitude differences across clusters despite differences in reading and language abilities.

It was expected that Cluster 1 would show poorest MMN amongst the four clusters while Cluster 4 would show the most robust presence of MMN across the two visits. The MMN amplitude results, however, were not as predicted. The detectability of MMN in individual participants did suggest a difference across clusters, however. Only 35 children (41%) across the four clusters showed MMN to both visits, however, most participants in Cluster 3 (79%) displayed a measurable MMN on visit 1, and 71% had MMN present across both visits. Language scores of children in Cluster 3 were the highest compared to other clusters and they had a relative strength in nonverbal intelligence.

Given the poor test-retest reliability of MMN and LDN in general and based on the current results, it would be strongly recommended that future studies consider either split-level reliability or undertake test-retest with both control and clinical populations. aSME is also recommended as an indicator of the quality of recordings. Another recommendation is to measure participants’ language ability as it seems that reliable MMN is related to good language skills, based on our Cluster 3 results and earlier research (Bishop, 2007). LDN showed poor detectability and significantly reduced amplitude when retested, therefore future research needs to explore paradigms and stimuli to improve LDN detection rates. Given that it is not certain why LDN reduced significantly in amplitude when children with listening difficulties were tested twice, perhaps the protocols could be more complex with similar but different stimuli, for instance use /da, ga, ba/ rather than just /da, ga/ and should also explore the effect of training effects. The results for different speech contrasts, with/without training may help to explain the effects of attention on LDN.

**5 CONCLUSION**

The present study shows high individual variability in MMN and LDN responses in children with listening difficulties and, overall, very poor detectability of LDN. LDN was reduced in amplitude at visit 2 for all children, irrespective of their strengths and difficulties. These findings question the feasibility of using LDN in future studies of children with listening difficulties due to poor test-retest reliability and poor detectability. The results suggest some possible short-term training effects on LDN or possibly changes in LDN due to lack of novelty of the stimuli at the second visit. For MMN, moderate test-retest reliability was found overall, Improvements in test protocol, perhaps through choice of stimuli, oddball paradigm parameters, or improved objective methods for response detection, are needed before MMN can be considered an appropriate tool for measuring the outcomes of auditory training in clinical populations (Wise et al., 2020). Individual children in Cluster 3 had better MMN reliability than other clusters. These MMN results support efforts to identify protocols for improving reliability of MMN in neurotypical and neurodiverse populations (Wise et al., 2020).

**Acknowledgements**

We thank for the children and their families who provided a lot of their time for this research.

**Conflict of interest statement**

The authors declare that the current research was carried out with no commercial or financial relationships and hence there are no conflicts of interest.

**Data Availability Statement**

The Auditory Evoked Response data relevant to this paper will be available on request from the corresponding author in a de-identified format after its publication.

**ORCID**

Mridula Sharma: [0000-0002-0448-6429](http://orcid.org/0000-0002-0448-6429)

Varghese Peter: 0000-0002-4007-507X

Suzanne C Purdy: [0000-0001-9978-8173](https://orcid.org/0000-0001-9978-8173)

**6 REFERENCES**

American Academy of Audiology. Clinical Practice Guidelines for the Diagnosis, Treatment and Management of Children and Adults with Central Auditory Processing Disorder 2010. https://audiology-web.s3.amazonaws.com/migrated/CAPD%20Guidelines%208-2010.pdf\_539952af956c79.73897613.pdf

Bishop DVM. Using mismatch negativity to study central auditory processing in developmental language and literacy impairments: where are we, and where should we be going? Psychol Bull 2007;133:651–72. https://doi.org/10.1037/0033-2909.133.4.651.

Bishop DVM, Hardiman MJ, Barry JG. Is auditory discrimination mature by middle childhood? A study using time-frequency analysis of mismatch responses from 7 years to adulthood. Dev Sci 2011;14:402–16. https://doi.org/10.1111/j.1467-7687.2010.00990.x.

Cheour M, Korpilahti P, Martynova O, Lang AH. Mismatch negativity and late discriminative negativity in investigating speech perception and learning in children and infants. Audiol Neurootol 2001;6:2–11. https://doi.org/10.1159/000046804.

Chermak GD, Tucker E, Seikel JA. Behavioral characteristics of auditory processing disorder and attention-deficit hyperactivity disorder: Predominantly inattentive type. J Am Acad Audiol 2002;13:332–8.

Cocquyt EM, Van Laeken H, van Mierlo P, De Letter M. Test–retest reliability of electroencephalographic and magnetoencephalographic measures elicited during language tasks: A literature review. Eur J Neurosci 2023; 57;1353-67. https://doi.org/10.1111/ejn.15948

Dalebout SD, Fox LG. Reliability of the mismatch negativity in the responses of individual listeners. J Am Acad Audiol 2001;12:245–53.

Davids N, Segers E, van den Brink D, Mitterer H, van Balkom H, Hagoort P, et al. The nature of auditory discrimination problems in children with specific language impairment: an MMN study. Neuropsychologia 2011;49:19–28. https://doi.org/10.1016/j.neuropsychologia.2010.11.001.

Dawes P, Bishop D. Auditory processing disorder in relation to developmental disorders of language, communication and attention: a review and critique. Int J Lang Commun Disord 2009;44:440–65. https://doi.org/10.1080/13682820902929073.

Delorme A, Makeig S. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J Neurosci Methods 2004;134:9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009.

Engström E, Kallioinen P, Lindgren M, Nakeva von Mentzer C, Sahlén B, Lyxell B, et al. Computer-assisted reading intervention for children with hearing impairment using cochlear implants: Effects on auditory event-related potentials and mismatch negativity. Int J Pediatr Otorhinolaryngol 2020;137:110229. https://doi.org/10.1016/j.ijporl.2020.110229.

Escera C, Alho K, Schröger E, Winkler I. Involuntary attention and distractibility as evaluated with event-related brain potentials. Audiol Neurootol 2000a;5:151–66. https://doi.org/10.1159/000013877.

Escera C, Yago E, Polo MD, Grau C. The individual replicability of mismatch negativity at short and long inter-stimulus intervals. Clin Neurophysiol 2000b;111:546–51. https://doi.org/10.1016/s1388-2457(99)00274-6.

Fitzgerald K, Todd J. Making sense of mismatch negativity. Front Psychiatry 2020;11:468. https://doi.org/10.3389/fpsyt.2020.00468.

Frey A, François C, Chobert J, Velay J-L, Habib M, Besson M. Music training positively influences the preattentive perception of voice onset time in children with dyslexia: A longitudinal study. Brain Sci 2019;9:E91. https://doi.org/10.3390/brainsci9040091.

Frodl-Bauch T, Kathmann N, Möller HJ, Hegerl U. Dipole localization and test-retest reliability of frequency and duration mismatch negativity generator processes. Brain Topogr 1997;10:3–8. https://doi.org/10.1023/a:1022214905452.

Guthrie D, Buchwald JS. Significance testing of difference potentials. Psychophysiology 1991;28:240–4. https://doi.org/10.1111/j.1469-8986.1991.tb00417.x.

Hakvoort B, van der Leij A, Maurits N, Maassen B, van Zuijen TL. Basic auditory processing is related to familial risk, not to reading fluency: An ERP study. Cortex 2015;63:90–103. https://doi.org/10.1016/j.cortex.2014.08.013.

Jerger J, Musiek F. Report of the consensus conference on the diagnosis of auditory processing disorders in school-aged children. J Am Acad Audiol 2000;11:467–74.

Kaan E, Barkley CM, Bao M, Wayland R. Thai lexical tone perception in native speakers of Thai, English and Mandarin Chinese: an event-related potentials training study. BMC Neurosci 2008;9:53. https://doi.org/10.1186/1471-2202-9-53.

Kathmann N, Frodl-Bauch T, Hegerl U. Stability of the mismatch negativity under different stimulus and attention conditions. Clin Neurophysiol 1999;110:317–23. https://doi.org/10.1016/s1388-2457(98)00011-x.

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of chiropractic medicine*, *15*(2), 155-163.

Koravand A, Jutras B, Lassonde M. Abnormalities in cortical auditory responses in children with central auditory processing disorder. Neuroscience 2017;346:135–48. https://doi.org/10.1016/j.neuroscience.2017.01.011.

Korpilahti P, Krause CM, Holopainen I, Lang AH. Early and late mismatch negativity elicited by words and speech-like stimuli in children. Brain Lang 2001;76:332–9. https://doi.org/10.1006/brln.2000.2426.

Korpilahti P, Lang H, Aaltonen O. Is there a late-latency mismatch negativity (MMN) component? Electroencephalogr Clin Neurophysiol 1995;95:P96. https://doi.org/10.1016/0013-4694(95)90016-G.

Kujala T, Leminen M. Low-level neural auditory discrimination dysfunctions in specific language impairment-A review on mismatch negativity findings. Dev Cogn Neurosci 2017;28:65–75. https://doi.org/10.1016/j.dcn.2017.10.005.

Leppänen PHT, Tóth D, Honbolygó F, Lohvansuu K, Hämäläinen JA, NEURODYS WP7 group, et al. Reproducibility of brain responses: High for speech perception, low for reading difficulties. Sci Rep 2019;9:8487. https://doi.org/10.1038/s41598-019-41992-7.

Leung JH, Purdy SC, Corballis PM. Improving emotion perception in children with autism spectrum disorder with computer-based training and hearing amplification. Brain Sci 2021;11:469. https://doi.org/10.3390/brainsci11040469.

Liasis A, Bamiou D-E, Campbell P, Sirimanna T, Boyd S, Towell A. Auditory event-related potentials in the assessment of auditory processing disorders: a pilot study. Neuropediatrics 2003;34:23–9. https://doi.org/10.1055/s-2003-38622.

Liu H-M, Chen Y, Tsao F-M. Developmental changes in mismatch responses to mandarin consonants and lexical tones from early to middle childhood. PLoS One 2014;9:e95587. https://doi.org/10.1371/journal.pone.0095587.

Lopez-Calderon J, Luck SJ. ERPLAB: an open-source toolbox for the analysis of event-related potentials. Front Hum Neurosci 2014;8. https://doi.org/10.3389/fnhum.2014.00213.

Lu S, Wayland R, Kaan E. Effects of production training and perception training on lexical tone perception--A behavioral and ERP study. Brain Res 2015;1624:28–44. https://doi.org/10.1016/j.brainres.2015.07.014.

Luck SJ, Stewart AX, Simmons AM, Rhemtulla M. Standardized measurement error: A universal metric of data quality for averaged event-related potentials. Psychophysiology 2021;58:e13793. https://doi.org/10.1111/psyp.13793.

Moore DR, Hunter LL. Auditory processing disorder (APD) in children: A marker of neurodevelopmental syndrome. Hearing, Balance Commun 2013;11:160–7. https://doi.org/10.3109/21695717.2013.821756.

Moreno S, Lee Y, Janus M, Bialystok E. Short-term second language and music training induces lasting functional brain changes in early childhood. Child Dev 2015;86:394–406. https://doi.org/10.1111/cdev.12297.

Näätänen R, Jacobsen T, Winkler I. Memory-based or afferent processes in mismatch negativity (MMN): a review of the evidence. Psychophysiology 2005;42:25–32. https://doi.org/10.1111/j.1469-8986.2005.00256.x.

Näätänen R, Pakarinen S, Rinne T, Takegata R. The mismatch negativity (MMN): towards the optimal paradigm. Clin Neurophysiol 2004;115:140–4. https://doi.org/10.1016/j.clinph.2003.04.001.

Pekkonen E, Rinne T, Näätänen R. Variability and replicability of the mismatch negativity. Electroencephalogr Clin Neurophysiol 1995;96:546–54. <https://doi.org/10.1016/0013-4694(95)00148-r>.

Petit, S., Badcock, N. A., Grootswagers, T., Rich, A. N., Brock, J., Nickels, L., Moerel, D., Dermody, N., Yau, S., Schmidt, E., & Woolgar, A. (2020). Toward an Individualized Neural Assessment of Receptive Language in Children. Journal of Speech, Language, and Hearing Research, 63(7), 2361–2385. https://doi.org/10.1044/2020\_jslhr-19-00313

‌

Picton TW, Alain C, Otten L, Ritter W, Achim A. Mismatch negativity: different water in the same river. Audiol Neurootol 2000;5:111–39. https://doi.org/10.1159/000013875.

Pihko E, Mickos A, Kujala T, Pihlgren A, Westman M, Alku P, et al. Group intervention changes brain activity in bilingual language-impaired children. Cereb Cortex 2007;17:849–58. https://doi.org/10.1093/cercor/bhk037.

Purdy SC, Sharma M, Morgan A. Measuring perceptions of classroom listening in typically developing children and children with auditory difficulties using the LIFE-UK questionnaire. J Am Acad Audiol 2018;29:656–67. https://doi.org/10.3766/jaaa.17053.

Putkinen V, Tervaniemi M, Saarikivi K, de Vent N, Huotilainen M. Investigating the effects of musical training on functional brain development with a novel Melodic MMN paradigm. Neurobiol Learn Mem 2014;110:8–15. https://doi.org/10.1016/j.nlm.2014.01.007.

Richardson U, Thomson JM, Scott SK, Goswami U. Auditory processing skills and phonological representation in dyslexic children. Dyslexia 2004;10:215–33. https://doi.org/10.1002/dys.276.

Roggia SM, Colares NT. Mismatch negativity in patients with (central) auditory processing disorders. Braz J Otorhinolaryngol 2008;74:705–11. https://doi.org/10.1016/S1808-8694(15)31380-X.

Ruhnau P, Herrmann B, Schröger E. Finding the right control: The mismatch negativity under investigation. Clin Neurophysiol 2012;123:507–12. https://doi.org/10.1016/j.clinph.2011.07.035.

Sharma M, Dhamani I, Leung J, Carlile S. Attention, memory, and auditory processing in 10- to 15-year-old children with listening difficulties. J Speech Lang Hear Res 2014;57:2308–21. https://doi.org/10.1044/2014\_JSLHR-H-13-0226.

Sharma M, Purdy SC, Humburg P. Cluster analyses reveals subgroups of children with suspected auditory processing disorders. Front Psychol 2019;10:2481. https://doi.org/10.3389/fpsyg.2019.02481.

Sharma M, Purdy SC, Kelly AS. Comorbidity of auditory processing, language, and reading disorders. J Speech Lang Hear Res 2009;52:706–22. https://doi.org/10.1044/1092-4388(2008/07-0226).

Sharma M, Purdy SC, Newall P, Wheldal K, Beaman R, Dillon H. Effects of identification technique, extraction method, and stimulus type on mismatch negativity in adults and children. J Am Acad Audiol, 2004;15; 616-32. https://doi.org/10.3766/jaaa.15.9.3

Sharma M, Purdy SC, Newall P, Wheldall K, Beaman R, Dillon H. Electrophysiological and behavioral evidence of auditory processing deficits in children with reading disorder. Clin Neurophysiol, 2006;117;1130-44. https://doi.org/10.1016/j.clinph.2006.02.001

Sharma M, Purdy SC, Newall P, Wheldall K, Beaman R, Dillon H. Electrophysiological and behavioral evidence of auditory processing deficits in children with reading disorder. Clin Neurophysiol 2006;117:1130–44. https://doi.org/10.1016/j.clinph.2006.02.001.

Sharma M, Purdy SC, Newall P, Wheldall K, Beaman R, Dillon H. Effects of identification technique, extraction method, and stimulus type on mismatch negativity in adults and children. J Am Acad Audiol 2004;15:616–32. https://doi.org/10.3766/jaaa.15.9.3.

Silman S, Silverman CA, Emmer MB. Central auditory processing disorders and reduced motivation: Three case studies. J Am Acad Audiol 2000;11:57–63.

Tervaniemi M, Lehtokoski A, Sinkkonen J, Virtanen J, Ilmoniemi RJ, Näätänen R. Test-retest reliability of mismatch negativity for duration, frequency and intensity changes. Clin Neurophysiol 1999;110:1388–93. https://doi.org/10.1016/s1388-2457(99)00108-x.

Tomlin D, Dillon H, Sharma M, Rance G. The impact of auditory processing and cognitive abilities in children. Ear Hear 2015;36:527–42. https://doi.org/10.1097/AUD.0000000000000172.

Tomlin D, Vandali A. Efficacy of a deficit specific auditory training program for remediation of temporal patterning deficits. Int J Audiol 2019;58:393–400. https://doi.org/10.1080/14992027.2019.1585586.

Uwer R, Albrecht R, von Suchodoletz W. Automatic processing of tones and speech stimuli in children with specific language impairment. Dev Med Child Neurol 2002;44:527–32. <https://doi.org/10.1017/s001216220100250x>.

Vergara-Martínez, M., Gomez, P., & Perea, M. (2020). Should I stay or should I go? An ERP analysis of two-choice versus go/no-go response procedures in lexical decision. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 46*(11), 2034–2048. https://doi.org/10.1037/xlm0000942

Wang J, Chen T, Jiao X, Liu K, Tong S, Sun J. Test-retest reliability of duration-related and frequency-related mismatch negativity. Neurophysiol Clin 2021;51:541–8. https://doi.org/10.1016/j.neucli.2021.10.004.

Wang Y, Ding N, Ahmar N, Xiang J, Poeppel D, Simon JZ. Sensitivity to temporal modulation rate and spectral bandwidth in the human auditory system: MEG evidence. J Neurophysiol 2012;107:2033–41. https://doi.org/10.1152/jn.00310.2011.

Weihing J, Chermak G, Musiek F. Auditory training for central auditory processing disorder. Semin Hear 2015;36:199–215. https://doi.org/10.1055/s-0035-1564458.

Wetzel N, Schröger E. On the development of auditory distraction: A review. Psych J 2014;3:72–91. https://doi.org/10.1002/pchj.49.

Wise KJ, Purdy SC, Maslin MR. Cortical evoked potentials and the mismatch negativity (MMN) – revisiting MMN’s potential and clinical utility. New Zealand Audiological Society Bulletin 2020;30:7.

Zachau S, Rinker T, Körner B, Kohls G, Maas V, Hennighausen K, et al. Extracting rules: early and late mismatch negativity to tone patterns. Neuroreport 2005;16:2015–9. https://doi.org/10.1097/00001756-200512190-00009.

**List of Tables and Figures**

**Table 1.** Cluster memberships, ages, means and standard deviations for performance on nonverbal intelligence (TONI-3), nonword reading and backward digit span, Phoneme Manipulation (QUIL subtest), Receptive language (CELF-IV) scores, attention and Frequency Pattern Test (FPT).

**Table 2:** The mean and standard deviation of accepted epochs for each cluster, stimulus type and visits

**Table 3.** MMN and LDN mean amplitudes (and standard deviation) in µV for the four clusters for the two visits at ~~Fz and~~ Cz

**Table 4.** ICC coefficient (95% confidence Intervals) for MMN and LDN at Cz for the two visits for each cluster

**Figure 1.** Grand averaged (n=85) control (red) and deviant waveforms (blue) at ~~Fz (top row)~~ ~~and~~ Cz for visit 1 (first left column) and visit 2 (middle column) as well as the deviant-control difference wave (third right column). The dark lines show the grand averaged waveform, and the shading depicts 95% confidence intervals.

**Figure 2.** Depiction of results for four clusters. Grand averaged difference waveforms for visit 1 (orange) and visit 2 (green) for the four clusters at ~~Fz and~~ Cz. The dark lines show the grand averaged waveform, and the shading depicts 95% confidence intervals. Boxplots representing the MMN amplitudes for visit 1 and 2 for ~~Fz and~~ Cz for 4 clusters.