The acceleration of spoken-word processing in children’s native-language acquisition: An ERP cohort study

Shiro Ojima, Hiroko Matsuba-Kurita, Naoko Nakamura, Hiroko Hagiwara

*Department of Language Sciences, Graduate School of Humanities, Tokyo Metropolitan University, Minami-Osawa, Hachioji, Tokyo 192-0397, Japan

+Research Institute of Science and Technology for Society (RISTEX), Japan Science and Technology Agency (JST), Nibancho, Chiyoda-ku, Tokyo 100-0004, Japan

© 2011 Elsevier Ltd. All rights reserved.

1. Introduction

Healthy adults can process their native languages at a remarkable speed. Language processing is sometimes viewed as a reflex-like phenomenon (Fodor, 1983). In normal continuous speech, adults produce and comprehend 120–200 words per minute (Crystal & House, 1990; Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967). Research using artificially accelerated speech shows that comprehension can be twice or three times faster than the normal rate (Dupoux & Green, 1997; Foulke & Sticht, 1969). These studies together indicate that adults can identify spoken words well within 200 ms on average given coherent contexts. Such rapid word identification is enabled by incremental processing of incoming speech signals. Listeners typically use phonetic information immediately as it becomes available and can identify words well before their endings (Marslen-Wilson, 1987). Efficient processing of word-initial information is thus essential for adults’ speedy spoken-word processing (Marslen-Wilson & Zwitserlood, 1989). Surprisingly, no conclusive evidence is available on how adult-like rapid spoken-word processing emerges during the course of children’s native-language acquisition (see below). The present study aims to provide new and reliable empirical evidence on this issue, using online measures of children’s brain activity.

Behavioral studies using reaction times provide evidence that incremental processing itself is in place even in 2-year-olds (Fernald, Pinto, Swingley, Weinberg, & McRoberts, 1998; Swingley, Pinto, & Fernald, 1999). However, behavioral measures requiring any kind of overt response seem to be ill-suited to study developmental changes in the speed of word processing. Generally, the cognitive processes involved in decision making and subsequent action execution have not fully developed in children (Bunge, Dudukovic, Thomason, Vaidya, & Gabrieli, 2002), probably due to the slow maturation of the prefrontal cortex critically involved in those processes (Sowell et al., 2003). Thus, age differences in reaction times to spoken words may result from processes involved in decision making and response execution, rather than language processes per se (Tyler & Marslen-Wilson, 1981). In fact, developmental decreases in auditory lexical-decision times can disappear after response-related age differences are taken into account (Edwards & Lahey, 1993). Children aged 5–8 seem to be slower by about 300 ms in responding to single spoken words than...
adults (e.g., Sekerina & Brooks, 2007; Valley & Metsala, 1990, note that the main focus of these studies is not the absolute speed of children’s word processing). However, such data cannot be taken straightforwardly as evidence that spoken-word processing accelerates by 300 ms from ages 5 to 8 onwards. Another major line of behavioral research focused on the difference between function words (closed-class words) and content words (open-class words) in children’s speech processing. One influential view is that while the processing of function words becomes adult-like at around age 10, that of content words is established much earlier (Friederici, 1983).

Here the event-related brain potential (ERP) technique, which has high temporal resolution, may contribute to the issue in a unique way. ERPs are real-time electrical reflections of on-going neural activities, obtained by averaging segments of electroencephalograms (EEGs) time-locked to a sensory or cognitive event (Rugg & Coles, 1995). They can visualize cortical processes of language without assigning the participant an explicit task (Bentin, Kutas, & Hillyard, 1995), and have the potential to disentangle the distinct sub-processes of word processing which occur before any conscious decision. ERPs can be safely acquired from children, and their usefulness in developmental language studies has already been established. For example, in 1-year-olds, ERPs already differ between words known to the children and those unknown (Mills, CoffeyCorina, & Neville, 1993; Mills, CoffeyCorina, & Neville, 1997; Mills et al., 2004). ERPs recorded in sentence contexts keep changing towards adult-like patterns until at least adolescence (Hahne, Eckstein, & Friederici, 2004; Holcomb, Coffey, & Neville, 1992).

In adult populations, ERPs have often been used to visualize incremental processing of spoken words. Manipulation of semantic contexts proved particularly useful in such attempts. Both spoken and written words which do not fit the preceding semantic context typically evoke an N400 response (Kutas & Hillyard, 1980; McCallum, Farmer, & Pocock, 1984). Semantic congruency effects have earlier onset in auditory than in visual ERPs (Holcomb & Neville, 1990, 1991). It is also long known that the onset of congruency effects in spoken words appears well before the acoustic offsets of those words (McCallum et al., 1984), which is compatible with the incremental nature of spoken-word processing. The onset of congruency effects sometimes manifests itself as an early negative ERP component that is distinct from the N400 (Connolly & Phillips, 1994; van den Brink, Brown, & Hagoort, 2001). This early negativity may be functionally separable from the N400 (Brown & Hagoort, 1993; Connolly & Phillips, 1994; Newman & Connolly, 2009; van den Brink et al., 2001). In particular, the earlier negativity, sometimes called the N200, may index the process of lexical selection, which is part of word processing proper (van den Brink et al., 2001; van den Brink & Hagoort, 2004), while the N400 itself has been claimed to reflect semantic integration, that is, the integration of the word into the context (Brown & Hagoort, 1993; for an opposing view, see Deacon, Hewitt, Yang, & Nagata, 2000). Another hypothesis calls the early negativity Phonological Mismatch/Mapping Negativity and regards it as reflecting initial phonological analysis preceding lexical selection (Connolly & Phillips, 1994; Newman & Connolly, 2009). Despite the differences in their functional interpretations of the early negativity, these studies together prove the usefulness of the ERP technique in illuminating the critical role of word-initial information in adults’ spoken-word processing. At the same time, they point to the possibility that developmental changes in the speed of word processing may appear at the onset of semantic congruency effects, rather than the peak of the N400.

In the present study, we aim to provide ERP evidence on how the acceleration of spoken-word processing takes place in the course of children’s language learning, focusing particularly on the onset of semantic congruency effects. Previous developmental ERP research manipulating semantic congruency focused on the N400 hypothesized to reflect semantic integration (Brown & Hagoort, 1993), and failed to report evidence on the acceleration of children’s spoken-word processing. In particular, ERPs collected from school-age children in a picture-word mismatch paradigm similar to ours contained no clear evidence for developmental changes in the speed of word processing (Byrne et al., 1999; Cummings, Ceponiene, Dick, Saygin, & Townsend, 2008). One view states that the semantic processing of single words is already established by 7 years of age (Cummings et al., 2008), which accords well with the claim that processing of content words is established well before age 10 (Friederici, 1983).

In general, the onset of an ERP response is difficult to analyze (Luck, 2005), and probably even more so in pediatric populations. Hence, to provide as reliable data as possible, we had 40 children in each analysis conducted. We adopted a cohort design, in which two age cohorts of primary-school children were followed up longitudinally. A cross-sectional design is currently more prevalent in developmental research, but direct evidence on development and/or learning can only be obtained from the same individual tested twice or more times longitudinally (e.g., Cheour et al., 1998; McLaughlin, Osterhout, & Kim, 2004). In the experiment, we used as critical stimuli highly frequent words that were most likely to have been acquired by primary-school children. This predicted that not only a typical N400 but also a late positive component (LPC) which appears last in mother-tongue acquisition (Juottonen, Revonsuo, & Lang, 1996) would be elicited by incongruous words, regardless of age. In the analyses of the onsets of congruency effects, we contrasted the following two hypotheses. One states that processing of spoken (content) words is well established by age 7, not only qualitatively (Cummings et al., 2008; Friederici, 1983) but also quantitatively. The alternative hypothesis is that the acceleration of single-word processing continues beyond age 7. This allows quantitative changes beyond age 7.

2. Methods

2.1. Participants

We studied two age cohorts, each of which consisted of 40 school-age Japanese children (total n = 80). They constitute a subset of the participants in our multi-purpose longitudinal neuroimaging project on children’s language development (Ojima, Nakamura, Matsuba-Kurita, Hoshino, & Hagiwara, 2011). All these 80 children had lived in Japan since birth, were right-handed (Oldfield, 1971), had no known disorder (psychological, neurological, auditory, or linguistic), had been born to a native Japanese-speaking mother, and were pupils at public elementary schools. The two age cohorts differed in the children’s ages. Children in the younger age cohort (hereafter, Age Cohort 1) were around 7 years of age, while those in the older age cohort (hereafter, Age Cohort 2) were around 9 years of age, at the time of their initial participation (Year 1). See Table 1 for more details. The two groups were roughly equal in the sex composition, handedness quotients, and socioeconomic status (measured by parents’ education level, types of occupation, income level, etc.). The data of 26 right-handed Japanese adults (13 men, mean age 21, range 18–29) were also analyzed. All children and their parents gave informed assent and consent, respectively. The adult participants gave informed consent. The ethics committee of Tokyo Metropolitan University approved the procedures.

2.2. Stimuli

Pre-recorded words pronounced by a female professional narrator were used as the critical experimental stimuli. They consisted of 80 basic-level Japanese content words (mean word length: 518 ms, SD 140 ms, mean log 10 of print frequency per million: 1.592 counts); more details can be found in a related report (Ojima et al., 2011). The results of a separate behavioral test conducted on 6-year-old preschoolers (n = 9) suggested that all these 80 words are highly likely to be comprehended by primary-school children. In a forced-choice picture-pointing task, the preschoolers listened to each of the 80 words and pointed to one picture that they thought matches the word, among four candidate pictures. The average accuracy rate of this task was 99.86%, suggesting that these 80 words are in the mental lexicon of even 6-year-old preschoolers. The participants in our ERP study are all primary-school children, and it is highly likely that all these children including the youngest participants (7-year-olds in Year 1) comprehended the meaning of all 80 words. In the ERP experiment, the playback of a spoken word was preceded by the appearance of
a colored illustration. The same recorded words were used in both Year 1 and Year 3, whereas different sets of illustrations were used in different years.

### 2.3. Procedure

We used a picture word mismatch paradigm, which presents participants with spoken words that are either semantically congruous or incongruous with the picture context (Byrne et al., 1999; Cummings et al., 2008; Friedrich & Friederici, 2004). This is a promising paradigm to study children’s single-word processing (1) because the direct comparison between congruous words and the same words presented in incongruous contexts can subtract out the effects of non-linguistic sensory processes, (2) because only one word is presented per trial so that the participant does not have to keep track of multiple words, (3) because working memory use is kept constant and due to the presence of the picture during word presentation, and (4) because previous research has already established that robust negative ERPs remiss of the N400 (Rutts & Hillyard, 1980) can be obtained even from 1-year-olds (Friedrich & Friederici, 2004, 2005). One trial sequence consisted of the appearance of a picture (colored illustration) at a visual angle of 10.61° × 10.61 degrees, the playback of a spoken word 1000 ms after picture onset (around 60 dB sound pressure level), and the disappearance of the picture 750 ms after word offset. This trial sequence repeated every 4000 ms. During the entire recording session, each of the 80 stimulus words was presented twice, once in a congruous context (congruous with the preceding picture) and once in an incongruous context. Incongruous words did not utilize rhyme with congruous words. Children took a rest every 80 trials. They were instructed to attend to the visual and auditory stimuli silently. No behavioral task was assigned. The congruent and incongruent stimuli were presented in a pseudorandom order. To ensure that each child was attending to the picture at every trial, an adult experimenter (Experimenter 1) accompanied the child in the EEG booth and controlled the progress of the stimulus presentation. Another experimenter (Experimenter 2) outside the EEG room monitored the child’s gaze (body movement, etc.) from a different angle via a surveillance camera. Experimenter 2 also controlled the progress of the stimulus presentation where necessary, and assisted Experimenter 1 by notifying him/her of any possible problem detected.

### 2.4. Data acquisition

In a sound-proof, electrically shielded room, EEGs were continuously recorded through 27 scalp electrodes arranged according to the Extended International 10-20 system (Sharbrough et al., 1991). NuAmp (Neuroscan, Inc., Texas) digitized EEGs at 500 Hz, using Fpz as the ground and the left earlobe as the online reference, with a bandpass filter of 0.1–100 Hz applied online. Vertical and horizontal electro-oculograms were recorded through the electrodes placed above and below the left eye and those placed besides the outer canthi. After applying an offline bandpass filter of 0.3–30 Hz, the continuous EEGs were segmented into epochs of 1700 ms consisting of 200 ms before word onset and 1500 ms after word onset. The scalp EEG channels were re-referenced to the average of both earlobes offline. Ocular artifacts on all scalp EEG channels were corrected by a method based on regression analysis (Semlitsch, Anderer, Schuster, & Presslich, 1986). Baselines were adjusted relative to the pre-stimulus 200 ms period. Epochs exceeding ±75 μV at the right electrode or ±150 μV at any scalp electrode were rejected automatically. The group means of the numbers of averaged trials are listed in Table 2. An analysis of variance (ANOVA) on the numbers of averaged trials revealed no significant effect involving Age Cohort (smallest p > .3). Although the main effect of Year was found to be significant (F(1, 156) = 10.68, p < .001), many enough trials were secured even in Year 1 (67–69 trials), to eliciting language-related ERPs reliably.

### 2.5. Analysis

First, to roughly grasp the onset of semantic congruency effects, we analyzed mean amplitudes during several time windows (TWs) before 400 ms, using a repeated-measures ANOVA. Mean amplitudes were obtained in the following three consecutive 50 ms TWs: 200–250, 250–300, and 300–350 ms. We analyzed the lateral electrodes as four regions: anterior (left: Fp1, F3, F7, right: Fp2, F4, F8), temporal (left: F5, T7, C5, right: F6, T8, C6), central (left: F1, C3, C1, right: F2, C4, C2), and posterior (left: P3, P7, O1, right: P4, P8, O2). Using a repeated-measures ANOVA, we analyzed the effects of Condition (congruous, incongruous) and its interactions with TW (3 levels), Hemisphere (left, right), and Region (anterior, temporal, central, posterior). The significance level was set at p = .05. Where the assumption of sphericity was violated, a correction procedure was applied (Huynh & Feldt, 1976).

The corrected p-value will be reported below, together with the epsilon (ε) and the uncorrected degrees of freedom. Significant interactions between Condition and Region were followed by a test of Condition effect at each level of Region. Year (1 vs. 3) was also included in the ANOVA, but as a between-subject factor, because the intra-individual correlations between years turned out to be small.

To further narrow down the onset of congruency effects, we ran paired t-tests comparing incongruous and congruous condition at each sampling point at each electrode, and sought for the earliest time point that meets the following two criteria: (1) the p-value is less than .05, and (2) the p-values for the subsequent 50 time points (corresponding to a period of 100 ms) are also less than .05 (Luck, 2005). The earliest point of divergence thus found will be reported as an estimate of the onset latency of congruency effects. The difference between Year 1 and Year 3 in the onset latency will be interpreted as an index of the acceleration of word processing. Note that we are interested in relative differences between years, not the absolute value of the latency. The onset latency estimated in the above way depends heavily on the number of participants; a larger n is more likely to lead to statistical significance at earlier latencies. The onset latencies estimated from groups of different sizes (e.g., children and adults in this study) should not be compared.

### 3. Results

#### 3.1. Children: an overview

The grand-average ERP waveforms comparing incongruous and congruous words in children are shown in Fig. 1. Regardless of group (Age Cohort 1 and Age Cohort 2) and year (Year 1 and Year 3), incongruous words clearly evoked a large N400-LPC complex. Because the LPC appears last in mother-tongue development (Juttenon et al., 1996), its appearance indicates that the mechanism of word processing at age 7 is qualitatively similar to that of adults. Thus we will focus on the quantitative aspects of word processing below. Fig. 2 presents a closer view of the early part of the waveforms. There appear to be two distinct negative peaks (N1 and N2) before the N400, for both congruous and incongruous words. This is in good accordance with the adults’ data reported by van den Brink et al. (2001). In Age Cohort 1 (around 7 years of age in Year 1), the ERPs for the two types of words begin to diverge at around the second negative peak (N200) in Year 1 (marked by an arrow in Fig. 2). However in Year 3 (2 years later), the divergence point seems to be around the P2 component (the first positive peak in the figure), which immediately precedes the second negative peak. Age Cohort 2 (around 9 years of age in Year 1) also has divergence points at around the P2 component. In this group, no large difference between years can be seen. Fig. 3 presents voltage maps of [Incongruous – Congruous] subtraction data in early TWs before 400 ms. The scalp distribution of the N400 effect seems to have a posterior dominance, with slight rightward lateralization, regardless of group and year.

#### 3.2. Children: mean amplitudes

In each age cohort, we compared congruous and incongruous conditions, using mean amplitudes during 200–250 ms, 250–300 ms, and 300–350 ms TWs. The results of repeated-measures ANOVAs are listed in Table 3.

#### 3.2.1. Age Cohort 1 (7-year-olds in Year 1)

In Age Cohort 1, a repeated-measures ANOVA involving Year as a factor revealed a statistically significant TW × Condition × Year interaction (F(2, 156) = 4.643, ε = .899, p = .014), suggesting that...
the temporal evolution of the Condition effect differed between Year 1 and Year 3. In addition, a significant TW × Condition interaction ($F(2, 156) = 38.64, \varepsilon = .899, p < .001$) and a significant TW × Condition × Region interaction ($F(6, 468) = 7.399, \varepsilon = .663, p < .001$) were found.

We further analyzed the effects of Condition in each year and in each TW separately. In Year 1, we found no statistically significant effects involving Condition in the first TW (200–250 ms, smallest $p = .175$). In the second TW (250–300 ms), the Condition main effect was significant (Table 3), marking the beginning of congruency effects (i.e., more negative potentials for incongruous than for congruous words). In the third TW (300–350 ms), the Condition main effect was highly significant (Table 3), as well as the Condition × Region interaction.

In Year 3, the Condition × Region interaction reached significance from the first TW (200–250 ms) (Table 3). ANOVAs separately conducted at each region in this TW revealed a significant Condition effect at the central ($F(1, 39) = 5.915, p = .020$) and posterior region ($F = 8.299, p = .006$), statistically supporting the appearance of the N400. In the second and third TW, both the Condition main effect and the Condition × Region interaction were highly significant (Table 3).

3.2.2. Age Cohort 2 (9-year-olds in Year 1)

In Age Cohort 2, no statistically significant interaction involving all of TW, Condition, and Year was found (smallest $p = .125$), suggesting that the temporal evolution of the Condition effects did not largely differ between Year 1 and Year 3 in the three TWs analyzed. The TW × Condition interaction ($F(2, 156) = 65.01, \varepsilon = .904, p < .001$) and the TW × Condition × Region interaction ($F(6, 468) = 7.4, \varepsilon = .576, p < .001$) were significant, as in Age Cohort 1. The three TWs and the 2 years were then analyzed separately, to further make sure that the temporal evolution of the Condition effect was similar across the 2 years. In line with this expectation, the Condition effect first became significant during the same TW across Year 1 and Year 3, as shown below.

In Year 1, the Condition × Region interaction as well as the Condition main effect was significant in the first TW (200–250 ms)

Table 2
Mean number of averaged trials (standard deviation in parentheses).

<table>
<thead>
<tr>
<th>Year 1</th>
<th>Year 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Congruous</td>
<td>Incongruous</td>
</tr>
<tr>
<td>Age Cohort 1</td>
<td>69.2 (10.6)</td>
</tr>
<tr>
<td>Age Cohort 2</td>
<td>68.2 (11.8)</td>
</tr>
</tbody>
</table>

Fig. 1. Children’s grand-average ERP waveforms at all electrodes. Negative is plotted upwards. ERP responses to spoken words that are semantically congruous (solid lines) or incongruous (dotted lines) with the picture contexts are shown separately for two age cohorts of children and for two years. Children in Age Cohort 1 were around 7 years of age in Year 1. Those in Age Cohort 2 were around 9 in Year 1.
Table 3
Results of repeated-measures ANOVAs on mean amplitudes.

<table>
<thead>
<tr>
<th></th>
<th>200–250 ms</th>
<th>250–300 ms</th>
<th>300–350 ms</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>d.f.</strong></td>
<td><strong>F</strong></td>
<td><strong>F</strong></td>
<td><strong>F</strong></td>
</tr>
<tr>
<td><strong>Age Cohort 1, Year 1</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Condition</td>
<td>1,39</td>
<td>&lt;1</td>
<td>6.005*</td>
</tr>
<tr>
<td>C x Hemisphere</td>
<td>1,39</td>
<td>&lt;1</td>
<td>2.917(*)</td>
</tr>
<tr>
<td>C x Region</td>
<td>3,117</td>
<td>1.044</td>
<td>1.936</td>
</tr>
<tr>
<td>C x H x R</td>
<td>3,117</td>
<td>1.732</td>
<td>2.876*</td>
</tr>
<tr>
<td><strong>Age Cohort 1, Year 3</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Condition</td>
<td>1,39</td>
<td>3.758(*)</td>
<td>8.640**</td>
</tr>
<tr>
<td>C x Hemisphere</td>
<td>1,39</td>
<td>3.124(*)</td>
<td>4.388*</td>
</tr>
<tr>
<td>C x Region</td>
<td>3,117</td>
<td>7.101**</td>
<td>6.470**</td>
</tr>
<tr>
<td>C x H x R</td>
<td>3,117</td>
<td>&lt;1</td>
<td>1.468</td>
</tr>
<tr>
<td><strong>Age Cohort 2, Year 1</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Condition</td>
<td>1,39</td>
<td>6.045*</td>
<td>24.80***</td>
</tr>
<tr>
<td>C x Hemisphere</td>
<td>1,39</td>
<td>2.776</td>
<td>&lt;1</td>
</tr>
<tr>
<td>C x Region</td>
<td>3,117</td>
<td>5.382**</td>
<td>7.701**</td>
</tr>
<tr>
<td>C x H x R</td>
<td>3,117</td>
<td>1.551</td>
<td>1.710</td>
</tr>
<tr>
<td><strong>Age Cohort 2, Year 3</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Condition</td>
<td>1,39</td>
<td>4.248*</td>
<td>25.56***</td>
</tr>
<tr>
<td>C x Hemisphere</td>
<td>1,39</td>
<td>&lt;1</td>
<td>1.718</td>
</tr>
<tr>
<td>C x Region</td>
<td>3,117</td>
<td>7.527**</td>
<td>10.58***</td>
</tr>
<tr>
<td>C x H x R</td>
<td>3,117</td>
<td>1.854</td>
<td></td>
</tr>
</tbody>
</table>

Notes: Statistically significant F-values are written in bold and accompanied by asterisks (*p < .05, **p < .01, ***p < .001). Marginally significant effects (.5 < p < .1) are marked by (*). Abbreviations: C = condition, H = hemisphere, R = region.

Fig. 2. Early parts of children’s grand-average ERP waveforms at a representative electrode (CP2) enlarged. Negative is plotted upwards. The points of divergence between congruous (solid lines) and incongruous words (dotted lines) are indicated by arrows. In Age Cohort 1 in Year 1, ERP to congruous and incongruous words begin to differ at around the second negative peak (N200). In the other instances, the points of divergence are at around the P2 component which immediately precedes the second negative peak.

Fig. 4. Presents estimates of onset latencies of semantic congruency effects at representative electrodes in children. Because the semantic congruency effects were slightly lateralized to the right hemisphere in children, data at right-hemisphere electrodes in the centroparietal regions are shown. Those at the midline electrodes are shown additionally. In Age Cohort 1, congruency effects began at around 260–290 ms in Year 1 and at around 180–240 ms in Year 3. The onset latency became earlier by about 45–95 ms in two years in this group, with the average across electrodes being about 70 ms. In Age Cohort 2, the onsets of congruency effects were around 195–235 ms in Year 1 and 200–230 in Year 3. The differences between years in this group are thus at a negligible level, in line with the mean amplitude analyses above.

3.3. Children: onset latencies

3.4. Adults

Similarly to children, adults clearly showed an N400 and an LPC in their grand-average ERP waveforms (Fig. 5A). A closer look at the waveforms (Fig. 5B) suggests that the ERP elicited by incongruous words begin to diverge from those elicited by congruous words at around 200 ms. This is not earlier than but actually quite similar to Age Cohort 2 and the Year-3 data of Age Cohort 1 (Fig. 2). The scalp distribution of the early congruency effect in adults (Fig. 5C) seems to lack a clear posterior dominance, and is very slightly lateralized to the left, rather than to the right hemisphere, unlike in children. Statistical analyses supported these observations. A repeated-measures ANOVA on mean amplitudes revealed that the Condition main effect was significant from the first TW (200–250 ms) (F(1, 25) = 7.587, p = .011), suggesting that semantic congruency effects were already present in this TW. The Condition effect was highly significant in the two subsequent TWs (250–300 ms: F = 32.06, p < .001). The Condition x Region interaction was only
Fig. 3. Children's voltage maps of semantic congruency effects at consecutive 50-ms time windows before 400 ms. Voltage differences obtained by [incongruous − congruous] subtraction are shown. Congruency effects are stronger towards centroparietal sites and are slightly lateralized towards the right hemisphere in all instances. Black triangles indicate the earliest time windows at which mean amplitudes statistically differed between congruous and incongruous words. Marginally significant in the third TW, $F(3, 75) = 3.012, \varepsilon = .700, p = .055$.

4. Discussion

Using the ERP technique, we traced the possible developmental changes in the speed of children's single-word processing. Semantic mismatches between picture contexts and spoken words elicited both an N400 and an LPC regardless of age. This indicates that children's mechanisms of spoken-word processing at age 7 are qualitatively similar to those of adults' (Cummings et al., 2008). The scalp distribution of the N400 also showed a posterior dominance consistently. The changes that can be expected from age 7 onward would thus be quantitative, rather than qualitative. Critically, in all groups we tested, congruency effects began to appear well before the acoustic offsets of the stimulus words (mean length: 518 ms), which is in good accordance with the incremental nature of spoken-word processing. In Age Cohort 1 (around 7 years of age initially), the onsets of semantic congruency effects became earlier by about 70 ms after 2 years. Such data support the hypothesis that the quantitative aspects of children's word processing are still changing after age 7. In Age Cohort 2 (around 9 years of age initially), no significant changes in the onset latency of congruency effects were observed after 2 years. Overall, congruency effects began at similar timings at or after age 9, including adults. In our longitudinal study design, ERP data were obtained repeatedly from the same two age cohorts. The differences between Year 1 and Year 3 in each cohort cannot be due to individual differences unrelated to age. One might argue that the difference of 70 ms found between Year 1 and Year 3 in Age Cohort 1 is due to habituation to the experimental procedure and/or experimental stimuli, but such an argument erroneously predicts that a similar difference between years would be found in Age Cohort 2 as well.

Overall, our ERP data are consistent with the view that the acceleration of spoken-word processing is in progress until somewhere between 7 and 9 years of age. Previous developmental ERP research...
on school-age or older children's word processing (Byrne et al., 1999; Cummings et al., 2008) focused on the N400, not on the onset of congruency effects, and failed to report developmental changes in the speed of word processing. In particular, Cummings et al. (2008) concluded that the semantic processing of single words is well established by 7 years of age. By focusing on the onsets of congruency effects, we extended these previous studies and provided ERP evidence showing that the temporal profile of single-word processing changes until somewhere between 7 and 9. We estimated that the onset latency of congruency effects became earlier by about 70 ms between ages 7 and 9. Given that only 200 ms or so is needed for adults to identify spoken words in contexts (Crystal & House, 1990; Dupoux & Green, 1997; Foulke & Sticht, 1969; Liberman et al., 1967), an initial advantage of 70 ms per one spoken word seems rather large. In older children and adults, semantic congruency effects first became statistically significant in the 200–250 ms TW. Acoustic information contained in the initial 200-ms segments in the word stimuli thus seems to play a key role. To further clarify how informative the initial acoustic information of our word stimuli is, we conducted a post-hoc behavioral experiment on 5 native Japanese-speaking adults in their twenties. They listened to each of the 80 critical words up to 100 ms, 150 ms, and 200 ms in a random order, and verbally reported the identities of the phonemes they could perceive. The average number of correctly identified phonemes was 1.52 (SD .040, range 1.48–1.59) for the initial 100-ms segment, 2.01 (SD .198, range 1.69–2.26) for the 150-ms segment, and 2.68 (SD .199, range 2.38–2.94) for the 200-ms segment. The corresponding numbers for primary-school children would be less than these. Thus the developmental changes we observed may be related to the processing of the initial one or two (or at most three) phonemes of the critical words. We expect that the first one or two phonemes are shared by hundreds of lexical items even in children’s small lexicon; searches of a Japanese lexical database return numbers like 2000 (Amano & Kondo, 2000). This indicates that the congruency effects we observed must have occurred well before acoustic information could narrow the lexical candidates down to one.

In the experiment, we used content words as critical stimuli; function words devoid of meaning are not suitable for semantic manipulation. Previous research highlighted children’s immature processing of function words (as opposed to content words) after age 7 (Friederici, 1983). It seems that both the processing of function words and that of content words undergo important developmental changes even after age 7. Our ERP data are not compatible with some behavioral studies which reported differences of about 300 ms between children up to 8 and adults in reaction times to spoken words (Sekerina & Brooks, 2007; Walley & Metsala, 1990; note that those differences are not the main focus of these studies). Such behavioral data probably reflect developmental changes in decision- or response-related cognitive processes, and not language processes alone (Edwards & Lahey, 1993; Tyler & Marslen-Wilson, 1981).

Previous research conducted in a picture–word mismatch paradigm report that the onset of the N400 effect (more negative ERPs for incongruous words) appears later in 1-year-olds by 300–400 ms than in adults. In 19-month-olds, the N400 effect first became significant in the 700–800 ms TW, whereas in adults, it was significant from 300 ms onwards (Friedrich & Friederici, 2004). In another study, 20-month-old (healthy) children showed significant N400 effects first in the 600–700 ms TW (no adults’ data reported; Torkildsen, Syversen, Simonsen, Moen, & Lindgren, 2007). Hence spoken-word processing seems to speed up gradually from infancy until at least age 7 and beyond. Previous infant studies also found early effects of congruency in a picture–word mismatch task, but these early effects reflect congruous words’ (rather than incongruous words’) more negative ERPs (Friedrich & Friederici, 2004). Major developmental changes do not seem to occur in this effect (Friedrich & Friederici, 2004, 2005). We noted only small signs of this effect in our children’s ERPs, which were, in many cases, statistically not significant. In our adults’ data, the effect was completely missing. Differences in stimulus properties...
may account for the different results across studies (Ojima et al., 2011).

One might be concerned that the two sets of 9-year-olds’ data are not completely the same, but we argue that they are similar enough in critical aspects. As shown in Table 3, the two datasets (Age Cohort 1, Year 3 and Age Cohort 2, Year 1) differ in the Condition main effect in the earliest TW (significant for Age Cohort 1, Year 3, but only marginally significant for Age Cohort 2, Year 1). However, what is more important is that both datasets show a significant Condition × Region interaction from this TW onward; follow-up analyses showed that the Condition effect is significant at, and stronger towards, posterior sites, in both datasets. As the main focus of the paper is the speed of spoken-word processing, it is critical that both datasets begin to show significant effects of Condition at posterior sites from the same TW. The minor difference between the two groups of 9-year-olds in the Condition main effect in the earliest TW could be due to individual differences. The presence of individual differences would be a problem in cross-sectional analyses. However, our main claim is drawn from the longitudinal data obtained repeatedly from the same individuals.

Our findings add to the growing body of literature which suggests that rich information processing of linguistic nature takes place before the peak of the N400 (Connolly & Phillips, 1994; Newman & Connolly, 2009; van den Brink & Hagoort, 2004; van den Brink et al., 2001; Van Petten, Coulson, Rubin, Plante, & Parks, 1999). The ERP waveforms in the present study contained a small negative peak that precedes and is distinct from the N400, whether or not the word matched the preceding picture in meaning. This negative peak is reminiscent of the N200 previously identified (van den Brink et al., 2001). The similarities in the waveform morphology are striking (for example, compare Fig. 3 of the present study and Fig. 2 of van den Brink et al., 2001). Originally, the semantic congruency effects on the N200 lacked a posterior scalp distribution typical of the N400 (van den Brink et al., 2001), but this observation was not replicated by a later study (van den Brink & Hagoort, 2004).

In our data too, congruency effects were consistently more pronounced towards posterior sites, regardless of the time windows (in line with van den Brink & Hagoort, 2004). Thus, the early congruency effects may share similar scalp distributions to the N400 (Van Petten et al., 1999). Functionally, the separation between the N400 and the early negative component (N200 or Phonological Mis-match/Mapping Negativity) may still be possible (Hagoort, 2008; Newman & Connolly, 2009). More precise functional characterization of the effects we observed must await future research.

As described in Section 2, the meaning of all the 80 words is highly likely to have been comprehended by all participants including the youngest children, because these words are in the lexicon of 6-year-old preschoolers. It is unlikely that the age effects we noted were due to differences in the percentage of words comprehended. A related issue is the age-of-acquisition (AOA) of words. Words are acquired at various ages. Behavioral studies consistently report that early AOA words are named faster than are late AOA words in picture-naming tasks, where the participant is asked to overtly articulate the names of objects in pictures (e.g., Barry, Hirsh, Johnston, & Williams, 2001). However, the effects of word AOA on behavioral data are not as uniform across different tasks as those of word frequency (Catling & Johnston, 2009). Critically, published reports of word AOA effects on ERPs are rather sparse. The only available studies we are aware of (Cuetos, Barbon, Urrutia, & Dominguez, 2009; Tainturier, Tamminen, & Thierry, 2005) both report that word AOA modulates the amplitudes of only late ERP responses after the peak of the N400. To our knowledge, no evidence exists that word AOA alters the onset latency of an ERP component or affects early ERP responses appearing before the peak of the N400. Also, behavioral effects of word AOA on naming latencies persist well into adulthood; significant interactions between word AOA and children’s ages are not often found (Anderson, 2008). Thus it is unlikely that word AOA can account for the age effects on ERPs we observed.

It remains ambiguous exactly what factors caused the developmental changes we observed in the speed of spoken-word processing. The maturation of the brain does not seem to be the sole factor, because language competence cannot be maintained without daily use (Pallier et al., 2003). It is likely that the interaction between experience and brain maturation contributed to the developmental changes we found. Linguistically normal children involve themselves in innumerable communicative interactions daily. Repetitive and sustained use of language may speed up the cortical processing of specific linguistic items and the processing mechanism in general. Children also start learning to read and write at around 6 years of age at public schools in many countries including Japan. One view posits that the acquisition of reading affects phonological processing by implementing fully segmental phonology (e.g., Morais, Bertelson, Cary, & Alegría, 1986), which may facilitate spoken-word processing. However, the presence of literate people who demonstrate fully segmental phonology challenges this view (Ventura, Kolinsky, Fernandes, Querido, & Morais, 2007). During childhood, the lateralization of language functions towards the left hemisphere gradually increases with age (e.g., Holland et al., 2001). More compact neural circuits in older children may enable more rapid information processing, although direct evidence demonstrating this relationship is lacking. We should also acknowledge the possibility that the causes of the developmental changes found here might not be specific to language processing. There might have been age-related changes in the way the participants spent the 1000 ms of picture preview prior to onset of the acoustic stimulus. Specifically, older children might have been more able or willing to generate a likely lexical candidate and perhaps covertly name it. However, the participants in our experiment were instructed to only attend to the stimuli and were not required to judge the semantic fitness of the picture–word pairs. It remains unclear whether children are willing to engage themselves in an additional task that is required at all.

A recent study has shown that highpass filtering beyond 0.1 Hz can attenuate late ERP components (Kappenman and Luck, 2010). We cannot take the onset latency of semantic congruency effects as something absolute, because high-pass filtering at 0.3 Hz may have distorted and attenuated the effects. However, because the same filter frequency was applied to all datasets, attenuation, if any, should have affected the two cohorts and the 2 years equally. Our claim for the acceleration of spoken-word processing is based on the relative difference between longitudinal datasets. We do not report onset latencies as absolute ones.

5. Conclusion

In conclusion, our data are consistent with the hypothesis that the acceleration of spoken-word processing continues beyond age 7. Basic linguistic items including frequently used words and core syntactic rules are acquired before age 7, but the online use of these pieces of knowledge seems to be refined in the following few years. During this period, the basis of adults’ speedy use of language is probably nurtured. Thus ages 7–9 seem to be an important period for the attainment of fully adult-like processing ability of the mother tongue.

Acknowledgements

We thank all the children and their families for taking part, Kenji Itoh for assisting audio recording, and Hideaki Koizumi for con-
tuous support. This study was supported by a Grant-in-Aid for the promotion of “Brain Science and Education, Type II” from the Research Institute of Science and Technology for Society, Japan Science and Technology Agency (RISTEX/JST) to H.H. Preparation of the revised manuscript was partly supported by a Grant-in-Aid for the Scientific Research (A) (No. 22242012) from the Japanese Ministry of Education, Culture, Sports, Science and Technology (MEXT) to H.H., and also by the Global COE program (No. D09) for the Centre for Advanced Research on Logic and Sensibility (CARLS) at Keio University, with which S.O. is now affiliated.

References


van den Brink, D., & Hagoort, P. (2004). The influence of semantic and syntactic context constraints on lexical selection and integration in spoken-word comprehen-

