International Journal of Psychophysiology 77 (2010) 195–204

Contents lists available at ScienceDirect

International Journal of Psychophysiology

journal homepage: www.elsevier.com/locate/ijpsycho

Multi-sensory learning and learning to read

Leo Blomert *, Dries Froyen

Department of Cognitive Neuroscience & Maastricht Brain Imaging Centre, Faculty of Psychology & Neuroscience, Maastricht University, The Netherlands

ARTICLE INFO

Article history:
Accepted 14 June 2010
Available online 17 June 2010

Keywords:
Letter-speech sound associations
Grapheme-phoneme correspondences
Multi-sensory processing
Audiovisual integration
Reading development

ABSTRACT

The basis of literacy acquisition in alphabetic orthographies is the learning of the associations between the letters and the corresponding speech sounds. In spite of this primacy in learning to read, there is only scarce knowledge on how this audiovisual integration process works and which mechanisms are involved. Recent electrophysiological studies of letter–speech sound processing have revealed that normally developing readers take years to automate these associations and dyslexic readers hardly exhibit automation of these associations. It is argued that the reason for this effortful learning may reside in the nature of the audiovisual process that is recruited for the integration of in principle arbitrarily linked elements. It is shown that letter–speech sound integration does not resemble the processes involved in the integration of natural audiovisual objects such as audiovisual speech. The automatic symmetrical recruitment of the assumedly uni-sensory visual and auditory cortices in audiovisual speech integration does not occur for letter and speech sound integration. It is also argued that letter–speech sound integration only partly resembles the integration of arbitrarily linked unfamiliar audiovisual objects. Letter–sound integration and artificial audiovisual objects share the necessity of a narrow time window for integration to occur. However, they differ from these artificial objects, because they constitute an integration of partly familiar elements which acquire meaning through the learning of an orthography. Although letter–speech sound pairs share similarities with audiovisual speech processing as well as with unfamiliar, arbitrary objects, it seems that letter–speech sound pairs develop into unique audiovisual objects that furthermore have to be processed in a unique way in order to enable fluent reading and thus very likely recruit other neurobiological learning mechanisms than the ones involved in learning natural or arbitrary unfamiliar audiovisual associations.

1. The role of multi-sensory learning in learning to read

The importance of written language in today's society can hardly be overestimated. It allows us to transcend our communication limits in space and time and proficient literacy has thus become a crucial marker of the quality of life (UNESCO, 2005). In the last decade neuroimaging studies have identified a brain region that shows specialisation for fast visual word recognition; i.e. the putative Visual Word form Area (Cohen et al., 2000) in the occipito-temporal cortex. Since fluency and automaticity are the most salient features of experienced reading it is indeed plausible that a neural network involved in visual object recognition has specialised for recognising visual letters and word forms (McCandliss et al., 2003). Fluency of reading however is an intriguing characteristic, since we need years of explicit instruction and practice before we start to exhibit any fluency in visual word recognition (Vaessen and Blomert, 2010) and persons with dyslexia may never attain a really fluent reading performance (Gabrieli, 2009). This long-lasting process contrasts sharply with the way we learn to master spoken language. Infants and young children start to pick up and develop the many complexities of spoken language without explicit instructions at a time when literacy instruction is still a potential event in a far future (e.g., Jusczyk, 1997). So we may ask the question: What makes learning to read so effortful?

One obvious answer to this question is that writing systems have only emerged fairly recently in evolution; i.e. a few thousand years ago and for the majority of people only a few hundred years ago (Rayner and Pollatsek, 1989). In contrast, spoken communication is an old habit and speech probably arose some 60,000 years ago (Lieberman, 2006). Therefore, it is very likely that our brains are evolutionarily prepared for speech, but not for learning to read. So, we may further ask: Which mechanisms enable the brain to learn to read?

A first hint may come from the fact that spoken language development not only precedes the learning of written language evolutionarily, but also ontogenetically: we speak and listen before we write and read. To find an answer we thus need a closer look at the beginnings of reading. The very first step in learning to read is establishing associations between letters and speech sounds (Frith, 1985; Marsh et al., 1981). In alphabetic languages written words are created out of a limited set of elements, i.e. letters. These letters, are purportedly representing their spoken counterparts, i.e. speech

* Corresponding author. Dept. Cognitive Neuroscience, Faculty of Psychology & Neuroscience, Maastricht University, P.O. Box 616, 6200 MD Maastricht, The Netherlands. Tel.: +31 43 3881949.
E-mail address: l.blomert@maastrichtuniversity.nl (L. Blomert).

0167-8760/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.ijpsycho.2010.06.025

sounds. Since children have already mastered spoken language to a considerable degree when they enter school, it has been suggested that written language builds on the spoken language system, particularly on the mechanisms for processing speech sounds (Liberman, 1973; Mattingly, 1972). By now it has been generally accepted that the ability to manipulate speech sounds and learning to read reciprocally influence each other during reading development (Perfetti et al., 1987). It has indeed been shown that the brain responses to phonological stimulation of healthy illiterates differed from the responses of normal readers (Castro-Caldas et al., 1998). This influence from learning to read on spoken phonological representations may occur because before learning to read the smallest elements of the spoken speech system are not isolated speech sounds or phonemes, but larger chunks of sound information which do not directly match onto the newly learned letters (Ziegler and Goswami, 2005). So, although the impression may be that the learning of an orthography directly connects to the already existing phonological representations of speech sounds, this may in effect consist of a re-shaping of the relevant spoken language elements, thus changing permanently the spoken language system. Although the mechanisms enabling these cross-modal influences are still unknown, we hypothesize that the formation of letter–sound associations very likely constitutes the vehicle via which learning to read changes spoken phonological representations. Considering the importance of letter–speech sound associations for learning to read, we may again rephrase our question to: How does the brain establish associations between letters and speech sounds?

Recent research has made clear that multi-sensory information processing is part and parcel of object perception and recognition in daily life, whereby the brain integrates the information from different modalities into a coherent percept (Ghazanfar and Schroeder, 2006). Neuroimaging research pointed to the superior temporal sulcus (STS) as an important brain area for audiovisual integration processes (e.g. Beauchamp et al., 2004; Calvert, 2001). Although this latter study mainly found activations for meaningful objects like animals and tools in this area, it was speculated that STS might also be instrumental for other kinds of audiovisual associations (see Hocking and Price, 2008 for a similar interpretation). The findings of a posterior temporal network for audiovisual speech processing (lip-reading), including the assumedly uni-sensory visual and auditory sensory cortices (e.g. Calvert et al., 1997, 1998) also sparked interest in letter–speech sound pairs as a special kind of audiovisual objects (Hashimoto and Sakai, 2004; Herdman et al., 2006; Raij et al., 2000). The interest for letters and their corresponding speech sounds partly stems from the fact that they are recent cultural inventions, sharply contrasting with e.g. audiovisual speech. The attraction also partly resides in the fact that letter–speech sound pairs are highly over-learned multi-sensory associations allowing to manipulate the congruency between the elements without activating higher order cognitive processes. Letter–speech sound pairs may thus be conceptualized as in principle arbitrary associations, which acquire meaning by learning a specific orthography. Although, as discussed above, one element of these associations seems already in place when learning to read starts, this is only partly true: the exact phonemes corresponding to letters or letter strings are not part of the neural and behavioural repertoire of spoken language before learning to read (Blomert and Willems, in press; Morais et al., 1979; Wimmer et al., 1991) and existing representations of speech sounds probably need a fundamental re-modelling to fit the requirements for adequate letter–speech sound associations. Thus, it is more adept to formulate that one element of the pair, i.e., phonemic speech sounds, are familiar in kind but not in type, when learning to read starts. These considerations point to the potentially ambiguous status of letter–sound pairs as audiovisual objects: Do letter–speech sound associations resemble natural audiovisual objects with known elements or do they resemble arbitrary associations between unfamiliar elements? And lastly, is it possible that the type of association (natural versus artificial) and the way these associations are used during reading imply different association mechanisms?

Before we review the findings on the neural correlates of letter–speech sound processing, we first need to clarify our concept of letter–sound associations. Although influential models of reading (Coltheart et al., 2001) and reading development (Ehri, 1995; Share, 1995) implemented a central role for the learning of letter–speech sound relations and its role in the learning of new words, any fundamental insights in the nature and workings of this association process are basically missing. This lack of basic research findings might, in part, be attributed to the widespread opinion that the associations between letters and speech sounds are mastered within a few months by most children in most alphabetic orthographies (e.g., Ziegler and Goswami, 2005). The assumed fast and easy learning process would by implication transfer the burden for explaining effortful and long-lasting fluent reading development to other processes. Therefore we want to emphasize the difference between the learning of letter-speech sound associations and “letter knowledge” or “letter–sound knowledge”, as “it is possible in principle for a child to know the modal pronunciations for all letters and still have not in place any notion that these sounds are parts of words” (Byrne and Fielding-Barnsley, 1989). Recent evidence indeed showed that even dyslexic children, who exhibited serious problems learning to read and learning letter–speech sound associations nevertheless showed full letter–knowledge mastery just like their normal reading peers towards the end of first grade (Blomert and Willems, in press). Furthermore, it was recently shown that the speed of letter-speech sound association processes systematically decreased over the full range of primary school grades without reaching a floor in sixth grade in normal readers (Blomert and Vaessen, 2009) suggesting an ongoing automation of these associations (Chein and Schneider, 2005). The salient differences in learning rate between letter knowledge and letter–speech sound associations suggest that an exploration of the type of audiovisual association, which is formed when learning letter–speech sound correspondences may provide key insights for understanding reading development.

2. Insights in letter–speech sound learning and integration

2.1. Magneto-encephalographic (MEG) insights in letter–speech sound associations

A rare early behavioural study revealed a first basic insight in letter–speech sound processing by investigating the influence of letter primes on the recognition of a speech sound in spoken syllables (Dijkstra et al., 1989). Subjects were asked to identify the vowel in a spoken syllable consisting of a consonant and a vowel. The target vowel was primed by a letter prime that was either congruent or incongruent with the target. The results showed clear decreases of response latencies if prime and target were congruent, thus indicating automatic cross-modal activations of speech sounds by letters.

It took another decade before a study appeared which investigated the neural correlates of the automatic audiovisual integration of letters and speech sounds with an emphasis on its temporal dynamics (Raij et al., 2000). This magneto-encephalographic (MEG) study reported no letter-specific cross-modal interaction effects in temporal sensory specific cortices. A first difference between the processing of letters (matching and non-matching) and non-letter control stimuli was recorded in the temporo-occipital–parietal junction around 225 ms after stimulus onset. And only the superior temporal sulci (STS) revealed strong interactions between letters and speech sounds 380–450 ms after stimulus onset. It is noteworthy that the interactions in the right hemisphere started almost 70 ms later than in the left STS, suggesting that the audiovisual integration process of letters and speech sounds mainly and first occurred in left STS. These results