Title	Spatio-temporal dynamics of automatic processing of phonological information in visual words
Author(s)	Wang, Xiao-Dong; Wu, Yin-Yuan; Liu, A.-Ping; Wang, Peng
Citation	Wang, X.-D., Wu, Y.-Y., Liu, A.-P., & Wang, P. (2013). Spatio-temporal dynamics of automatic processing of phonological information in visual words. Scientific Reports, 3, 3485.
Date	2013
URL	http://hdl.handle.net/10220/19328
Rights	

SCIENTIFIC REPORTS

OPEN

Correspondence and
requests for materials
should be addressed to
X.-D.W. (xd-wang@
ntu.edu.sg)

Spatio-temporal dynamics of automatic processing of phonological information in visual words

Xiao-Dong Wang[1], Yin-Yuan Wu[2], A.-Ping Liu[3] & Peng Wang[3]

[1]Division of Psychology, School of Humanities and Social Sciences, Nanyang Technological University, Singapore, [2]The F. M. Kirby Neurobiology Center, Boston Children's Hospital, Department of Neurology, Harvard Medical School, Boston, USA, [3]CAS Key Laboratory of Ion Beam Bioengineering, Institute of Technical Biology and Agriculture Engineering, Chinese Academy of Sciences, Hefei, China.

Sensory-specific cortices appear to be sensitive to information from another modality. Here we investigate whether the human brain automatically extracts the phonological information in visual words in early visual processing. We continuously presented native Chinese speakers peripherally with Chinese homophone characters in an oddball paradigm, while they performed a visual detection task presented in the centre of the visual field. We found the lexical tone phonology embedded in the characters is processed automatically by the brain of native speakers, as revealed by whole-head electrical recordings of the mismatch negativity (MMN). Source solution further revealed the MMN involved the neural activations from the visual cortex to the auditory cortex (130–460 ms). The spatial-temporal dynamics indicate a visual-auditory interaction in the early, automatic processing of phonological information in visual words.

It is interesting and counterintuitive that sensory-specific cortices appear to be sensitive to information from another modality[1,2]. Audiovisual speech processing which involves multisensory integration sites including low level auditory and visual sensory regions[3,4] is ecologically important in daily communications.

Empirical evidence from behavioural and neuroimaging studies demonstrated that in audiovisual speech processing, human auditory cortex responds to visual information and vice versa. A typical example of audio-visual speech processing is the integration of a written and a spoken language. Learning the correspondence between a spoken word and its written form is a prerequisite for the development of reading and writing skills[5]. Behavioural studies of written word recognition demonstrated the role of auditory phonology in visual word processing[6–10]. Some recent neuroimaging studies showed that phonological processing during visual word recognition is governed mainly by the left brain hemisphere including left lateralized supramarginal gyrus and left lateralized inferior frontal cortex, as well as bilateral superior temporal gyri[11–14]. In event-related potential (ERP) studies of phonological processing in written words, the majority of research focused on N400 or even later stage components[15–18]. A few other ERP studies pointed to relatively early phonological effects on the early components such as P2 and N2[19–22].

The issue of whether or not phonological information embedded in written words can be automatically extracted by the brain and rapidly stored in visual sensory memory has remained virtually unexplored. To address this issue, an effective approach to isolate the brain response component contributed by the cognitive processing at an early, preattentive stage with a high temporal resolution is essential. To this end, the mismatch negativity (MMN) can be a potential tool. MMN has been suggested an index of early, automatic encoding of the changes of regularities in both of the auditory[23–25] and visual modality[26]. In audition, MMN can be elicited by deviants differing in sound intensity, duration, frequency and abstract regularities in a sound stream. In vision, MMN is named visual MMN (vMMN), can be elicited by deviants differing in spatial frequency, colour, line orientation, shape and even abstract sequential regularities in a visual stimuli stream. It has been verified a memory-based change detection mechanism that underlies MMN in both the auditory[27] and visual[28,29] modalities, and MMN can be a potential tool to study the function of the auditory and visual sensory memory.

To use MMN as a tool to investigate the automatic extraction of auditory phonological information in a stream of written words, an important matter is to select visual words with the same phonological information to form the word stream. To this point, Chinese homophones can be ideal materials. In Mandarin Chinese, there are arbitrary mappings between the orthographic and phonological forms, homophones are broadly used, and one
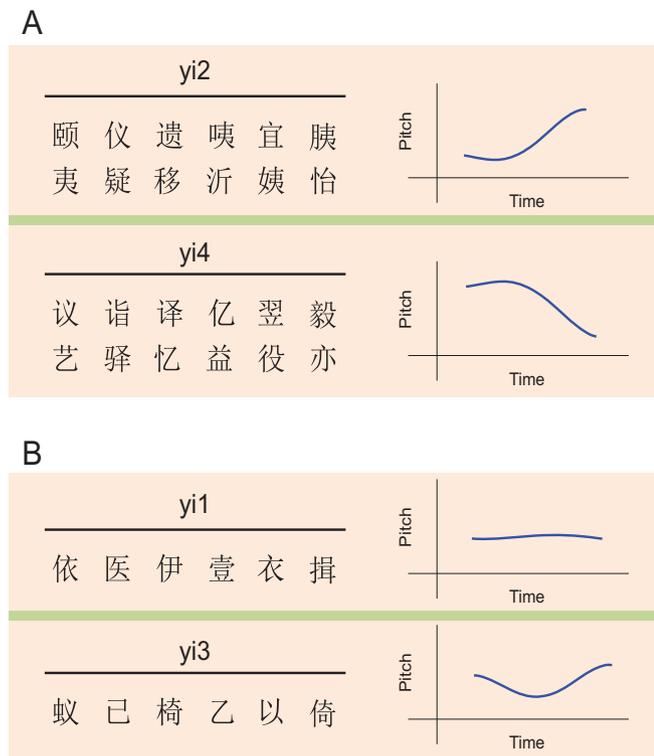
A

B

Figure 1 | Visual stimuli of Chinese characters. Four homophone pools were constructed. A: The homophone pool [yi2] with the rising tone and pool [yi4] with the falling tone. Each contained 12 characters. B: The homophone pool [yi1] with the flat tone and pool [yi3] with the dipping tone. Each contained 6 characters.

spoken syllable usually associates to multiple written characters, and even vice versa. Furthermore, unlike alphabetic orthographies which always differ in the length of words (number of letters), the Chinese homophones are all single block characters, though they may differ

in the number of strokes. These features make Chinese characters ideal materials to be used in a visual oddball paradigm. In our previous study, we presented subjects with Chinese characters in the centre of the visual field and found a rapid processing of phonological information, as revealed by the vMMN[30]. As the visual characters were not presented perifoveally, the vMMN recorded in that study may not be fully automatic. The issue of whether or not phonological information in visual words is automatically processed by the brain needs to be further explored.

In the present study, participants were presented peripherally with visual stimuli of four Chinese homophone characters (Fig. 1 and 2). The four characters that made up a stimuli trial were randomly selected from a pool of Chinese homophones. Participants were required to focus on the central fixation cross and make a speeded button-press when the cross changed in the length of arms. We continuously presented advanced Mandarin Chinese speakers with stimuli trials in an oddball sequence. Each trial consisted of four Chinese homophones (with the same phonological information). The four characters appeared together on the upper-left, upper-right, lower-left and lower-right part of the monitor. The standard stimuli and deviant stimuli were comprised of four homophones picked randomly from separate homophone pools (Fig. 1) which differed in the lexical tone. The experimental paradigm was adapted from Stefanics' recent studies of vMMN[31,32]. We hypothesize that if the MMN-like response can be elicited in this oddball paradigm using homophones, then phonological information is extracted automatically by the brain of natives, indicating that the visual sensory memory has been already sensitive to phonological information at an early, preattentive processing stage.

Results

Behavioural response to the occasional changes in the fixation cross showed a high hit rate (>96% for all subjects). The reaction times for block 1 (standard_yi2, deviant_yi4), block 2 (standard_yi4, deviant_yi2), block 3 (standard_yi1, deviant_yi3) and block 4 (standard_yi3, deviant_yi1) were 443 ms, 436 ms, 445 ms and 441 ms respectively. The mean reaction time for all conditions was 441 ± 12.2 ms (s.e.m.).
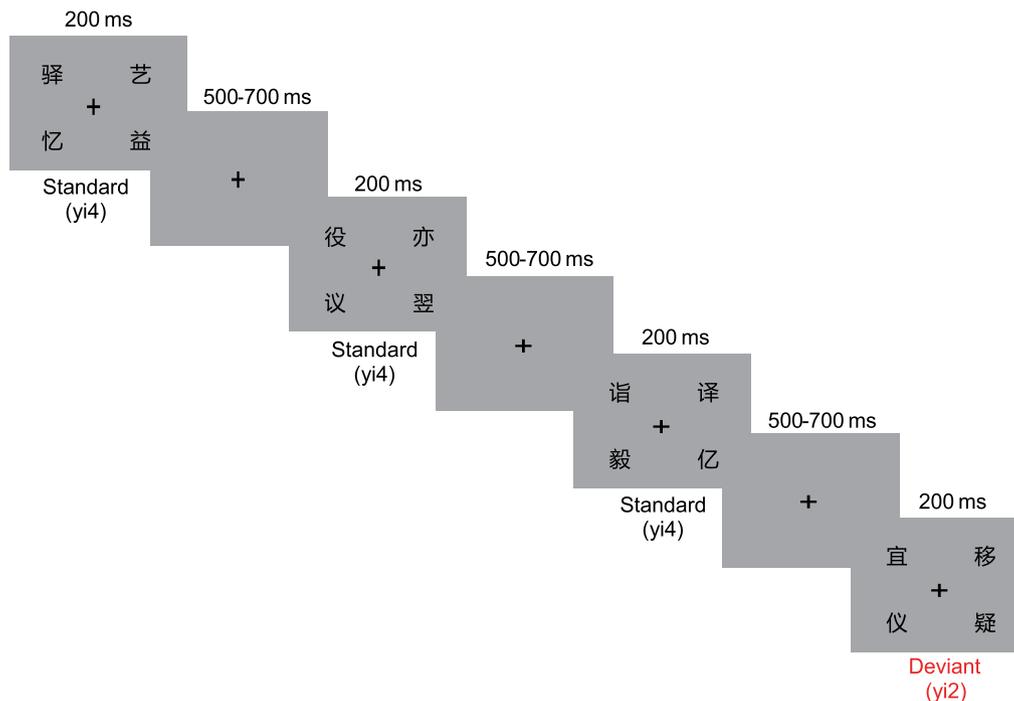

Figure 2 | Experimental design and a sample of the oddball block. Participants' task was a speeded button-press to the changes of the central fixation cross. Characters were presented peripherally. Both the standard and deviant stimuli consisted of four characters randomly picked from separate homophone pools.

A robust P1-N170-P2 complex was evoked by both standard and deviant stimuli at the parieto-occipital and fronto-central sites (Fig. 3 Left). The N170 response to the standard stimuli in the left occipital region was stronger than that in the right occipital region ($t(11) = -2.599$, $p = 0.025$, two tailed) by comparing the mean amplitude of PO3, PO5, O1 and PO7 electrodes in the left parieto-occipital region to that of PO4, PO6, O2 and PO8 electrodes in the right parieto-occipital region. However, there was no significant difference of N170 response between the left fronto-central region (mean amplitude of F3, F5, FC3 and FC5) and the right fronto-central region (mean amplitude of F4, F6, FC4 and FC6).

There was an obvious negative deflection in the deviant ERP compared to the standard ERP (Fig. 3 and Supplementary Fig. S2). Based on the visual inspection of the difference waveforms and the topographic maps, three time windows (130–190 ms, 190–390 ms and 390–460 ms) were defined to calculate the mean amplitude of ERP. Two representative electrodes in the parieto-occipital region

(Pz and Oz) and two in the fronto-central region (Fz and FCz) were chosen to illustrate the ERP waveforms (Fig. 3 Left). To analyze the significance of the difference between the ERP in response to the standard stimuli and deviant stimuli, a three-way ANOVA (repeated measures) was conducted with stimuli condition (standard and deviant), site (fronto-central and parieto-occipital) and time window (130–190 ms, 190–390 ms and 390–460 ms) the factors. The results showed a three-way interaction ($F(2, 10) = 11.861$, $\varepsilon = 0.685$, $p = 0.002$) and an interaction between stimuli condition and time window ($F(2, 10) = 4.055$, $\varepsilon = 0.980$, $p = 0.033$), as well as a main effect of stimuli condition ($F(1, 11) = 7.065$, $\varepsilon = 1.000$, $p = 0.022$). We further conducted a two-way ANOVA (repeated measures) with stimuli condition and time window the factors at each level of the factor site, i.e., fronto-central and parieto-occipital. In the fronto-central sites, there was an interaction between stimuli condition and time window ($F(2, 10) = 6.253$, $\varepsilon = 0.782$, $p = 0.013$), and a main effect of stimuli condition ($F(1, 11) = 5.931$, $\varepsilon = 1.000$, $p = 0.033$). In

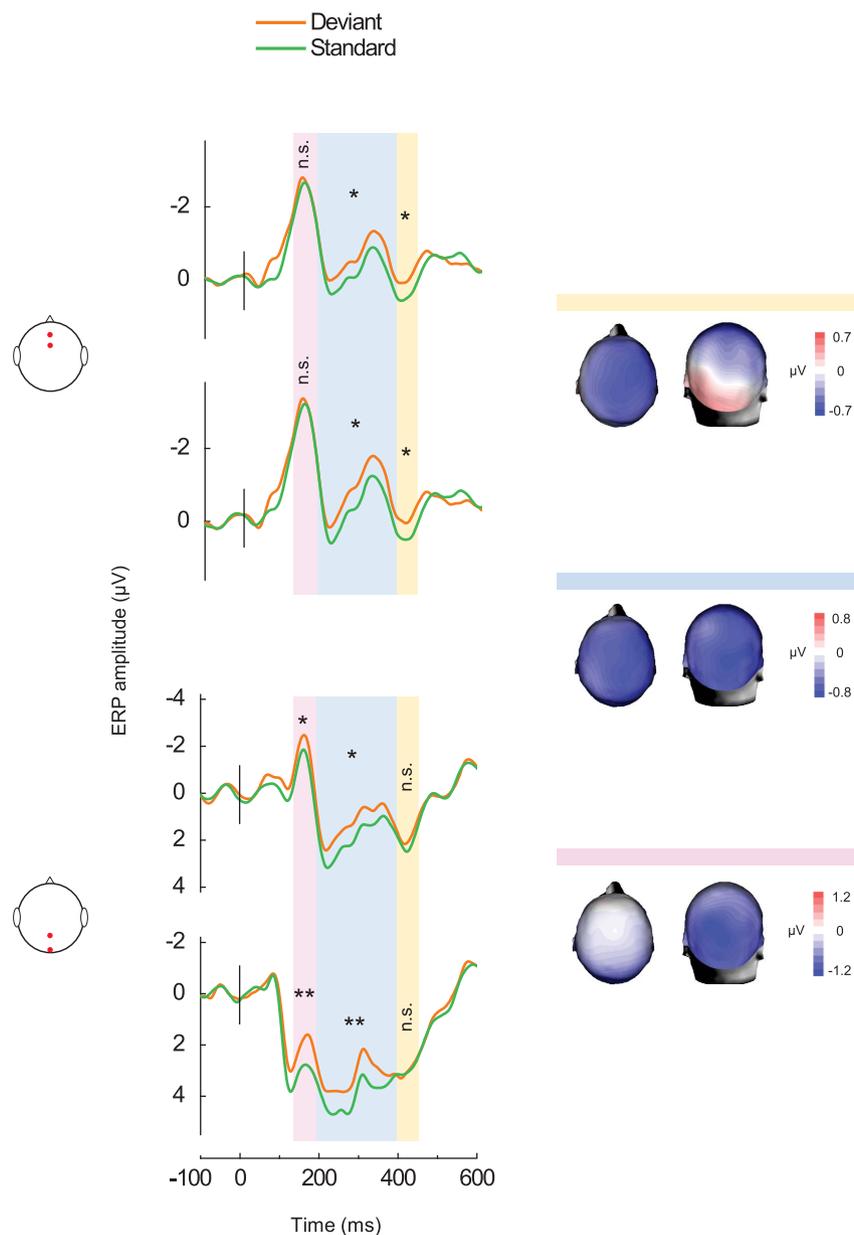

Figure 3 | Grand averaged event-related potential (ERP) responses to the standard and deviant stimuli. Left: Grand averaged ERPs for standard stimuli and deviant stimuli at fronto-central sites Fz and FCz (top), and parieto-occipital sites Pz and Oz (bottom). Right: Scalp topographic maps of the difference wave at peak latencies of the 130–190 ms, 190–390 ms and 390–460 ms time intervals.

the parieto-occipital sites, there was an interaction between stimuli condition and time window ($F(2, 10) = 11.984$, $\varepsilon = 0.641$, $p = 0.002$), and a main effect of stimuli condition ($F(1, 11) = 7.811$, $\varepsilon = 1.000$, $p = 0.017$). We further conducted a paired t-test of the mean amplitude of ERPs in response to the deviants and standards at Fz, FCz, Pz and Oz for each time window. In the 130–190 ms time window, Pz ($t(11) = -2.284$, $p = 0.043$), Oz ($t(11) = -3.227$, $p = 0.008$). In the 190–390 ms range, Pz ($t(11) = -2.545$, $p = 0.027$), Oz ($t(11) = -3.110$, $p = 0.01$), Fz ($t(11) = -2.323$, $p = 0.040$), FCz ($t(11) = -2.457$, $p = 0.032$). In the 390–460 ms range, Fz ($t(11) = -2.425$, $p = 0.034$), FCz ($t(11) = -2.332$, $p = 0.040$). Consistent findings can be found in Supplementary S2.

As the average word occurrence frequency of homophone pool [yi3] was much higher than that of the other three homophone pools (620 vs. 35, 46 and 44 per million), we also measured whether the word occurrence frequency had an effect on the MMN elicitation. Since the paradigm design was block 1 (standard_yi2, deviant_yi4), block 2 (standard_yi4, deviant_yi2), block 3 (standard_yi1, deviant_yi3) and block 4 (standard_yi3, deviant_yi1), we compared the MMN mean amplitude (standard vs. deviant) elicited in block 1 and 2 to that elicited in block 3 and 4. The results showed no significant differences between the MMN mean amplitude elicited in block (1 + 2) and that elicited in block (3 + 4). The waveforms and the statistics were shown in the supplementary section (see Supplementary Fig. S1 and Table S1b online).

The scalp topographic maps of MMN were also constructed with the three time ranges. At the 130–190 ms time window, the MMN topographic maps showed only the property of vMMN response (e.g., a parieto-occipital distribution). At the 190–390 ms time window, the MMN topographic maps showed properties of both the auditory MMN and vMMN (e.g., fronto-central as well as parieto-occipital activation). At the 390–460 ms time window, the MMN topographic maps showed only the property of auditory MMN (e.g., a fronto-central distribution) (Fig. 3 Right).

The distinct spatial distributions of MMN over temporal stages were further supported by a source solution using swLORETA method, which evaluated the intracranial source of the MMN elicited. We performed the swLORETA source analysis at each time window (Fig. 4 and Table 1). Results showed that at the 130–190 ms time window, the current source was most prominent in the occipital lobe (BA 19). At the 190–390 ms time window, the current sources were prominent in occipital, limbic, temporal, and frontal lobe (BA 19, 30, 21 and 10). At the 390–460 ms window, the current source was prominent in temporal and frontal lobe (BA 21, 6 and 9).

Discussion

We investigated the ERP correlates of automatic processing of lexical tone phonology in written characters outside the focus of participants' attention. We found that the lexical tone phonology embedded in Chinese characters was automatically extracted and rapidly stored in visual sensory memory, as revealed by the MMN. Furthermore, using a source solution of swLORETA, we found that the MMN in response to the contrast of the lexical tone phonology in written characters involved the neural activations from the visual cortex to the auditory cortex. The spatial-temporal dynamic patterns indicate a visual-auditory interaction in the early processing of phonological information in a visual word stream. We suggest that this visual-auditory interactive processing in the natives is due to the activation of long-term memory of the lexical tones embedded in the written characters.

To address the issue of automatic extraction of phonological information in visual words, vMMN, an index of early and automatic encoding of changes in the visual modality[33], is an ideal tool. However, without proper written word materials, this issue can hardly be addressed. For instance, vMMN can be elicited if two visual words from an alphabetic language are presented in an oddball sequence, but

the vMMN will contain the contributions of the contrasts in orthographies, meanings and phonologies. It is necessary to remove the contamination from orthographic or semantic contrasts, so as to investigate the phonological processing in written words exclusively. A potential solution is to use several different homophones as stimuli in an oddball sequence, in which some homophones are presented as standard stimuli and the other homophones are presented as deviant stimuli. The standard stimuli and deviant stimuli differ in some specific phonological features. The contrasts in orthographies or meanings will not be extracted because the visual words are constantly changing. To this end, Chinese homophone characters can be ideal materials. A Chinese character is the smallest unit of meaning. In the mapping between the orthographic and phonological representations, unlike most alphabetic languages in which there is an intimate correspondence between the written and phonological forms, there is no letter to sound correspondence in logographic Chinese. The mapping between visual and phonological forms is relatively arbitrary in Chinese: One spoken syllable is often associated with multiple written characters (even vice versa) and in Chinese, homophones are broadly used. This unique property of Chinese characters provides us an opportunity to explore the automatic extraction of phonological information in a changing stream of written words.

In the present study, we chose Chinese homophones to create four homophone pools: [yi1], [yi2], [yi3] and [yi4] (Fig. 1). Characters of the four homophone pools were matched for average stroke numbers. The average occurrence frequency of the characters of the homophone pools [yi1], [yi2] and [yi4] was also matched, except homophone pool [yi3] which had a higher average occurrence frequency. In this study, we did not find a frequency effect in the MMN results (see Supplementary Fig. S1 and Table S1b online). This is in line with some previous findings which showed that MMN response to auditory words was not sensitive to the occurrence frequency of the words[34]. For instance, in a study on the MMN elicited by Finnish auditory words, a word elicited larger MMN amplitude than an acoustically matched pseudo-word. When compared this MMN enhancement (MMN for word – MMN for pseudo-word) for a word with a higher occurrence frequency to that for a word with a lower occurrence frequency, there was no significant difference[35]. With respect to this issue, previous studies demonstrated more positive-going ERPs with higher word frequencies, which was observed for later ERP components with longer latencies, such as P300, N400 and P600[36,37]. However, there was also evidence showed that the word occurrence frequency modulated the amplitude of auditory MMN[38]. For the MMN response to visual words, to the best of our knowledge, there is no evidence that word occurrence frequency has an effect on the vMMN elicitation.

Our results verified the automatic extraction of phonological information in visual words. In our previous study, we presented Chinese homophones in the centre of the visual field and found a rapid processing of lexical tone phonology in Chinese characters[30]. In that study, participants were required to detect the characters with a different colour. Although the task was irrelevant to orthography, semantic or phonological processing, participants might still possible to have expectancy for the repetition of the homophone, and hence the vMMN we found in that study might not be fully automatic. In the present study, we used four-character stimuli and the four characters were presented on the periphery of the visual field. This paradigm excluded the possibility of subjects' expectancy for the repetition of homophones and the elicitation of vMMN in the present study will further verify the automatic processing of phonological information embedded in written words. The vMMN we recorded spanned from a wide time range (130 to 460 ms), and we proposed that the automatic and preattentive processes might only happen in the earlier time range (e.g., 130–190 ms).

It is well known that N170 is a face-sensitive ERP component[39]. Visual stimuli of Chinese characters also evoked the N170 brain
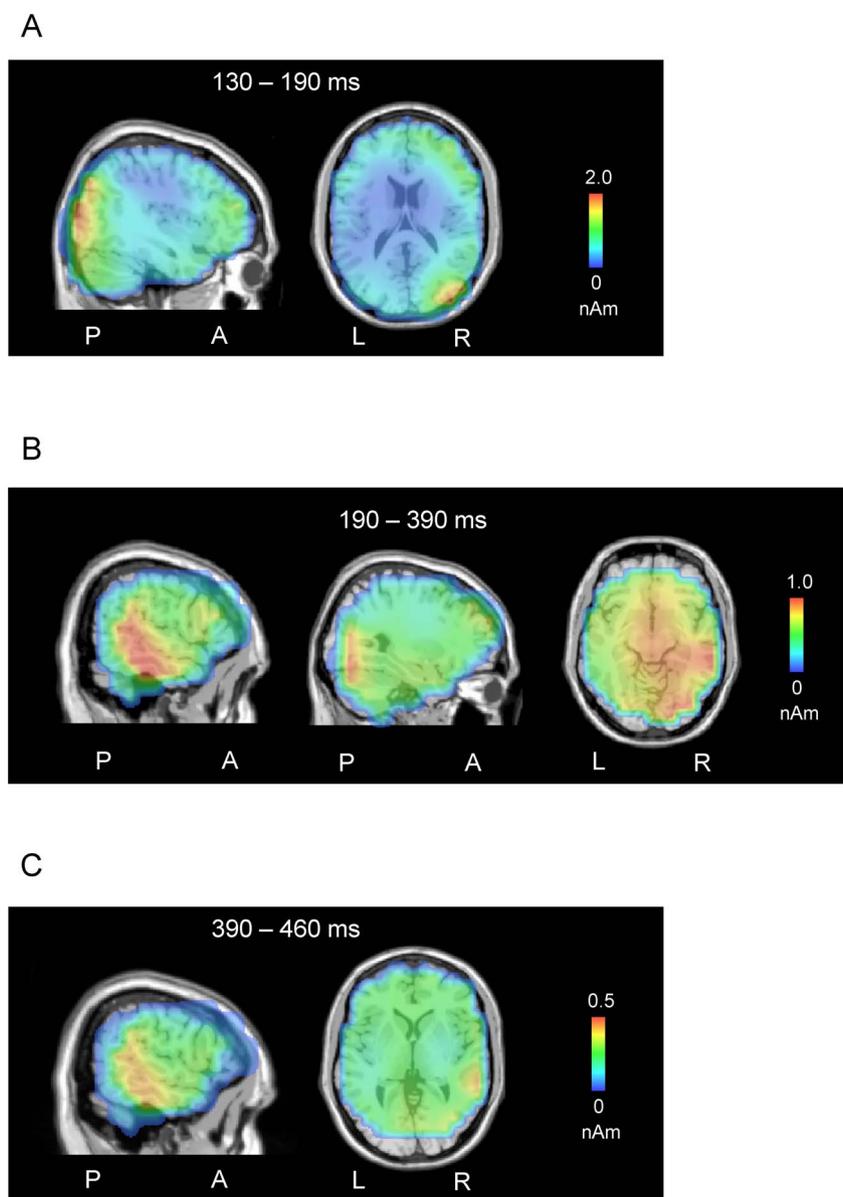
Figure 4 | Intracranial active sources explaining the MMN according to the swLORETA analysis. The same intracranial activation area was shown in a sagittal and an axial view in the 130–190 ms, 190–390 ms and 390–460 ms time intervals.

response[40]. In the present study, the visual stimuli evoked a robust N170 component, and it was stronger in the left occipital region than in the right, and this reflects the neural change resulting from extensive experience and expertise with a script[41]. Our results are in line with current literature arguing that N170 response to written words is typically left lateralized in skilled readers[42].

Whilst having been correlated with neural automatic change detection in vision and short-term sensory memory[28], vMMN has rarely been investigated in the context of long-term neural representations, but see[30]. Our results further verified the use of vMMN as a tool to explore the long-term memory of lexical tone phonology, which depends on the long-term learning and experience. In the auditory domain, numerous studies revealed the effects of long-term memory and experience on the early auditory processing of speech sounds[35,43,44]. Our results indicate that this is also true in the visual domain. Besides, the MMN (Fig. 3 and 4) showed properties of both vMMN and auditory MMN. We suggest that the phonological processing in written words may not be executed and completed within a single modality, but rather involves the functional connections of both visual and auditory pathways. vMMN has been considered to

be a potential tool for cognitive dysfunction and learning[45–47]. In this sense, the experimental procedure deployed in the present study could be applied to test the effect of learning of the correspondences between spoken words and their written forms, which is a prerequisite for the development of reading and writing skills.

In summary, our results showed the automatic extraction of lexical tone phonology in visual characters as revealed by vMMN, suggesting that visual sensory memory is sensitive to phonological information at an early processing stage. Our results indicate the rapid and automatic activation of long-term memory of phonological information embedded in written characters, and will help to understand the neural mechanisms underlying our remarkable capacity of visual cortex in encoding of phonological information. Further research needs to be done to determine the corresponding cortico-cortical functional connections for this effect.

Methods

Subjects. Twenty-two subjects with no history of neurological or psychiatric impairment were recruited. All subjects were right-handed according to an assessment with the Edinburgh Handedness Inventory[48]. They were advanced adult

Table 1 | Talairach coordinates (in mm) corresponding to intracranial generators explaining the scalp MMN recorded in response to deviance in the 130–190 ms, 190–390 ms and 390–460 ms time windows. Magnitude in (Am E-10); BA = Brodmann areas

Magnitude	T-x	T-y	T-z	Hemisphere	Lobe	Gyrus	BA
				130–190 ms			
40.33	40	−73	25	R	O	Superior Occipital Gyrus	19
37.51	−33	−91	21	L	O	Middle Occipital Gyrus	19
33.46	−47	−87	−5	L	O	Inferior Occipital Gyrus	19
27.31	18	0	58	R	F	Sub-Gyral	6
25.40	23	−44	32	R	Limbic	Cingulate Gyrus	31
23.65	23	46	5	R	F	Superior Frontal Gyrus	10
14.81	27	−79	36	R	P	Precuneus	19
14.59	55	2	−5	R	T	Superior Temporal Gyrus	22
12.79	−33	−46	53	L	P	Superior Parietal Lobule	7
Magnitude	T-x	T-y	T-z	Hemisphere	Lobe	Gyrus	BA
				190–390 ms			
43.59	28	−70	12	R	Limbic	Posterior Cingulate	30
42.55	58	−22	−4	R	T	Middle Temporal Gyrus	21
37.96	20	50	14	R	F	Superior Frontal Gyrus	10
37.68	−50	−69	−4	L	O	Middle Occipital Gyrus	19
29.08	38	−23	49	R	P	Postcentral Gyrus	3
21.40	−10	−1	66	L	F	Superior Frontal Gyrus	6
19.23	19	49	−15	R	F	Superior Frontal Gyrus	11
16.24	21	−76	39	R	P	Precuneus	7
Magnitude	T-x	T-y	T-z	Hemisphere	Lobe	Gyrus	BA
				390–460 ms			
15.05	60	−22	−5	R	T	Middle Temporal Gyrus	21
11.32	50	2	43	R	F	Precentral Gyrus	6
10.24	9	−20	57	R	F	Medial Frontal Gyrus	6
7.99	−9	60	38	L	F	Superior Frontal Gyrus	9

native Mandarin Chinese speakers with good reading and writing skills. Twelve subjects (7 females, age range = 22–30 years) participated in the main experiment and ten subjects (6 females, age range = 23–28 years) participated in the supplementary experiment (Supplementary Information S2). The experimental protocol was approved by the institutional review board of the Institute of Technical Biology and Agriculture Engineering of Chinese Academy of Sciences. All participants had normal or corrected-to-normal vision and provided written informed consent.

Stimuli and Procedure. Each stimuli trial used in this study consisted of four Chinese homophones differing in orthography and meaning. The characters were appearing together on the upper-left, upper-right, lower-left and lower-right positions on an LCD monitor. By virtue of the unique property of Chinese homophones, we created four homophone pools (Fig. 1), which differed in the lexical tones. The phonologies of the characters in the four pools were Chinese syllable [yi] with the flat tone, rising tone, dipping tone and falling tone, respectively (yi1, yi2, yi3 and yi4). The stimuli trials were presented in an oddball sequence: standard stimuli trials were presented with a probability of 85%, while deviant stimuli trials 15%. The standard stimuli and deviant stimuli differed in the tone. There were four blocks and the block order was fully random between subjects: Block 1 (standard_yi2, deviant_yi4), block 2 (standard_yi4, deviant_yi2), block 3 (standard_yi1, deviant_yi3) and block 4 (standard_yi3, deviant_yi1). In this way, for each pair of the homophone pools (Fig. 1), the standard and deviant stimuli swapped (deviant-standard-reverse paradigm). In each block, a total of 1100 stimuli trials were presented. Each block was presented once and the block order was randomized across subjects. The homophone pools [yi2] and [yi4] both contained 12 homophones and the homophone pools [yi1] and [yi3] both contained 6 homophones. The experimental design was illustrated in Fig. 2. Each character stimuli subtending 5.7° visual angle horizontally and 7.7° vertically were presented on an LCD monitor on a middle grey background at a viewing distance of 50 cm. The centre-to-centre distance between horizontal and vertical characters were 20 cm and 13.5 cm. The average occurrence frequency of the characters of the four homophone pools was 35, 46, 44 and 620 per million for [yi2], [yi4], [yi1] and [yi3] respectively[49]. The average stroke numbers of the homophone pool [yi2], [yi4], [yi1] and [yi3] were 9.3, 7.3, 8.5 and 6.5 respectively. Each stimuli trial was presented pseudorandomized with a duration of 200 ms and an interstimulus interval of 500–700 ms. Two stimuli never appeared in immediate succession. Between two deviant stimuli, there were at least three standard stimuli. Participants were instructed to focus on the fixation cross presented in the centre of the visual field and press a button as accurately and quickly as possible when the cross

changed in the length of its arms. The fixation cross became wider or longer randomly, with an average frequency of 12 changes in every 100 trials. Participants were told that the characters presented peripherally were irrelevant. The hit rate and reaction time were recorded.

Data recording and analysis. EEG was recorded (SynAmps amplifier, NeuroScan) with a cap carrying 64 Ag/AgCl electrodes placed at standard locations covering the whole scalp (the extended international 10–20 system). Signals were filtered on-line with a low-pass of 100 Hz and sampled at a rate of 500 Hz. The reference electrode was attached to the tip of the nose, and the ground electrode was placed on the forehead. Vertical electrooculography (EOG) was recorded using bipolar channel placed above and below the left eye, and horizontal EOG was recorded using bipolar channel placed lateral to the outer canthi of both eyes. Electrode impedances were kept < 5 k Ohm. The recording data were filtered off-line between 1 and 25 Hz (24 dB/octave) with a finite impulse response filter. Epochs were set at 800 ms in length, starting 100 ms before the onset of stimulus. Epochs obtained from continuous data were rejected when fluctuations in potential values exceeded ± 100 µV at any channel except the EOG channels. The ERPs evoked by standard and deviant stimuli were calculated by averaging individual trials (excluding standards that immediately followed a deviant and the ERP response to trials where the cross changed in the length of arms). MMN, a difference waveform, was derived by subtracting the ERP response to the stimuli presented as standards in one block from the ERP response to the same stimuli presented as deviants in another block. By this means, deviant and standard ERP responses to physically identical stimuli (characters from the same homophone pool) were compared (deviant_yi1 vs. standard_yi1, deviant_yi2 vs. standard_yi2, deviant_yi3 vs. standard_yi3 and deviant_yi4 vs. standard_yi4).

The mean amplitudes of the ERPs were analyzed using a three-way ANOVA (repeated measures) with stimuli condition (standard and deviant), site (fronto-central and parieto-occipital) and time window (130–190 ms, 190–390 ms and 390–460 ms) as factors. Because the main goal of this study is to examine whether the difference between the ERPs in response to the standards and the deviants is significant, only the interactions and main effects concerning stimuli condition will be further analyzed. The Greenhouse-Geisser adjustment was applied, and corrected p values were reported along with uncorrected degrees of freedom.

In order to localize the source of the MMN in the brain, we performed the Low-Resolution Electromagnetic Tomography (LORETA) at a time window where the MMN response was most prominent. LORETA is a tomographic technique to the inverse EEG problem and helps find the solution consistent with the EEG\ERP scalp

topographic distribution. Here, we used an improved version of standardized weighted LORETA, so-called swLORETA. swLORETA incorporated a singular value decomposition based the lead field weighting method[50], and was performed on the group average data (switched to average reference) and evaluated statistically significant electromagnetic dipoles ($p < 0.05$). The grid spacing which is the distance between two calculation points was set up to 15. We ran the solution with a realistic boundary element model (BEM), obtaining from a standard MRI data set (a result of averaging 27 T1 weighted MRI acquisitions from a single male subject).

1. Sharma, J., Angelucci, A. & Sur, M. Induction of visual orientation modules in auditory cortex. *Nature* **404**, 841–847 (2000).
2. von Melchner, L., Pallas, S. L. & Sur, M. Visual behaviour mediated by retinal projections directed to the auditory pathway. *Nature* **404**, 871–876 (2000).
3. Calvert, G. A., Campbell, R. & Brammer, M. J. Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Curr Biol* **10**, 649–657 (2000).
4. Macaluso, E., George, N., Dolan, R., Spence, C. & Driver, J. Spatial and temporal factors during processing of audiovisual speech: a PET study. *Neuroimage* **21**, 725–732 (2004).
5. Ehri, L. C. Development of sight word reading: phases and findings. *The Science of Reading: A Handbook*, eds Snowling, M. J. & Hulme, C. (Oxford: Blackwell Publishing), 135–145 (2005).
6. Berent, I. Phonological priming in the lexical decision task: regularity effects are not necessary evidence for assembly. *J Exp Psychol Hum Percept Perform* **23**, 1727–1742 (1997).
7. Ferrand, L. & Grainger, J. Phonology and orthography in visual word recognition: evidence from masked non-word priming. *Q J Exp Psychol A* **45**, 353–372 (1992).
8. Pexman, P. M., Lupker, S. J. & Jared, D. Homophone effects in lexical decision. *J Exp Psychol Learn Mem Cogn* **27**, 139–156 (2001).
9. Whatmough, C., Arguin, M. & Bub, D. Cross-modal priming evidence for phonology-to-orthography activation in visual word recognition. *Brain Lang* **66**, 275–293 (1999).
10. Ziegler, J. C., Ferrand, L., Jacobs, A. M., Rey, A. & Grainger, J. Visual and phonological codes in letter and word recognition: evidence from incremental priming. *Q J Exp Psychol A* **53**, 671–692 (2000).
11. Bles, M. & Jansma, B. M. Phonological processing of ignored distractor pictures, an fMRI investigation. *BMC Neurosci* **9**, 20 (2008).
12. Hickok, G. *et al.* A functional magnetic resonance imaging study of the role of left posterior superior temporal gyrus in speech production: implications for the explanation of conduction aphasia. *Neurosci Lett* **287**, 156–160 (2000).
13. Hickok, G. & Poeppel, D. The cortical organization of speech processing. *Nat Rev Neurosci* **8**, 393–402 (2007).
14. Stoeckel, C., Gough, P. M., Watkins, K. E. & Devlin, J. T. Supramarginal gyrus involvement in visual word recognition. *Cortex* **45**, 1091–1096 (2009).
15. Bentin, S., Mouchetant-Rostaing, Y., Giard, M. H., Echallier, J. F. & Pernier, J. ERP manifestations of processing printed words at different psycholinguistic levels: time course and scalp distribution. *J Cogn Neurosci* **11**, 235–260 (1999).
16. Newman, R. L. & Connolly, J. F. Determining the role of phonology in silent reading using event-related brain potentials. *Brain Res Cogn Brain Res* **21**, 94–105 (2004).
17. Proverbio, A. M., Vecchi, L. & Zani, A. From orthography to phonetics: ERP measures of grapheme-to-phoneme conversion mechanisms in reading. *J Cogn Neurosci* **16**, 301–317 (2004).
18. Rugg, M. D. Event-related potentials and the phonological processing of words and non-words. *Neuropsychologia* **22**, 435–443 (1984).
19. Barnea, A. & Breznitz, Z. Phonological and orthographic processing of Hebrew words: electrophysiological aspects. *J Genet Psychol* **159**, 492–504 (1998).
20. Meng, X., Jian, J., Shu, H., Tian, X. & Zhou, X. ERP correlates of the development of orthographical and phonological processing during Chinese sentence reading. *Brain Res* **1219**, 91–102 (2008).
21. Niznikiewicz, M. & Squires, N. K. Phonological processing and the role of strategy in silent reading: behavioral and electrophysiological evidence. *Brain Lang* **52**, 342–364 (1996).
22. Ren, G. Q., Liu, Y. & Han, Y. C. Phonological activation in chinese reading: an event-related potential study using low-resolution electromagnetic tomography. *Neuroscience* **164**, 1623–1631 (2009).
23. Paavilainen, P. The mismatch-negativity (MMN) component of the auditory event-related potential to violations of abstract regularities: a review. *Int J Psychophysiol* **88**, 109–123 (2013).
24. Wang, X. D., Gu, F., He, K., Chen, L. H. & Chen, L. Preattentive extraction of abstract auditory rules in speech sound stream: a mismatch negativity study using lexical tones. *PLoS One* **7**, e30027 (2012).
25. Wang, X. D., Wang, M. & Chen, L. Hemispheric lateralization for early auditory processing of lexical tones: Dependence on pitch level and pitch contour. *Neuropsychologia* **51**, 2238–2244 (2013).
26. Czigler, I. Visual mismatch negativity: violation of nonattended environmental regularities. *Journal of Psychophysiology* **21**, 224–230 (2007).
27. Naatanen, R., Paavilainen, P., Rinne, T. & Alho, K. The mismatch negativity (MMN) in basic research of central auditory processing: a review. *Clin Neurophysiol* **118**, 2544–2590 (2007).
28. Czigler, I., Balazs, L. & Winkler, I. Memory-based detection of task-irrelevant visual changes. *Psychophysiology* **39**, 869–873 (2002).
29. Maekawa, T., Tobimatsu, S., Ogata, K., Onitsuka, T. & Kanba, S. Preattentive visual change detection as reflected by the mismatch negativity (MMN)--evidence for a memory-based process. *Neurosci Res* **65**, 107–112 (2009).
30. Wang, X. D., Liu, A. P., Wu, Y. Y. & Wang, P. Rapid extraction of lexical tone phonology in Chinese characters: a visual mismatch negativity study. *PLoS One* **8**, e56778 (2013).
31. Stefanics, G., Csukly, G., Komlosi, S., Czobor, P. & Czigler, I. Processing of unattended facial emotions: a visual mismatch negativity study. *Neuroimage* **59**, 3042–3049 (2012).
32. Stefanics, G., Kimura, M. & Czigler, I. Visual mismatch negativity reveals automatic detection of sequential regularity violation. *Front Hum Neurosci* **5**, 46 (2011).
33. Czigler, I. Visual Mismatch Negativity and Categorization. *Brain Topogr*, *doi:10.1007/s10548-013-0316-8* (2013).
34. Pulvermuller, F., Shtyrov, Y., Kujala, T. & Naatanen, R. Word-specific cortical activity as revealed by the mismatch negativity. *Psychophysiology* **41**, 106–112 (2004).
35. Pulvermuller, F. *et al.* Memory traces for words as revealed by the mismatch negativity. *Neuroimage* **14**, 607–616 (2001).
36. Rugg, M. D. Event-related brain potentials dissociate repetition effects of high- and low-frequency words. *Mem Cognit* **18**, 367–379 (1990).
37. Polich, J. & Donchin, E. P300 and the word frequency effect. *Electroencephalogr Clin Neurophysiol* **70**, 33–45 (1988).
38. Alexandrov, A. A., Boricheva, D. O., Pulvermuller, F. & Shtyrov, Y. Strength of word-specific neural memory traces assessed electrophysiologically. *PLoS One* **6**, e22999 (2011).
39. Rossion, B., Joyce, C. A., Cottrell, G. W. & Tarr, M. J. Early lateralization and orientation tuning for face, word, and object processing in the visual cortex. *Neuroimage* **20**, 1609–1624 (2003).
40. Lin, S. E. *et al.* Left-lateralized N170 response to unpronounceable pseudo but not false Chinese characters-the key role of orthography. *Neuroscience* **190**, 200–206 (2011).
41. Brem, S. *et al.* Neurophysiological signs of rapidly emerging visual expertise for symbol strings. *Neuroreport* **16**, 45–48 (2005).
42. Dehaene, S. Electrophysiological evidence for category-specific word processing in the normal human brain. *Neuroreport* **6**, 2153–2157 (1995).
43. Naatanen, R. *et al.* Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature* **385**, 432–434 (1997).
44. Shtyrov, Y. & Pulvermuller, F. Neurophysiological evidence of memory traces for words in the human brain. *Neuroreport* **13**, 521–525 (2002).
45. Chang, Y. *et al.* Dysfunction of preattentive visual information processing among patients with major depressive disorder. *Biol Psychiatry* **69**, 742–747 (2011).
46. Kimura, M., Schroger, E. & Czigler, I. Visual mismatch negativity and its importance in visual cognitive sciences. *Neuroreport* **22**, 669–673 (2011).
47. Qiu, X. *et al.* Impairment in processing visual information at the pre-attentive stage in patients with a major depressive disorder: a visual mismatch negativity study. *Neurosci Lett* **491**, 53–57 (2011).
48. Oldfield, R. C. The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* **9**, 97–113 (1971).
49. Liu, Y. *et al.* Modern Chinese frequency dictionary. (China Astronautic Publishing House, Beijing, 1990).
50. Palmero-Soler, E., Dolan, K., Hadamschek, V. & Tass, P. A. swLORETA: a novel approach to robust source localization and synchronization tomography. *Phys Med Biol* **52**, 1783–1800 (2007).

Acknowledgments

Author contributions

X.W. conceived the initial idea of this research. X.W. and Y.W. designed methods and experiments. Y.W. and A.P. interpreted the results and wrote the paper. X.W. and P.W. performed data acquisition, Y.W. co-worked on source localization analysis. All authors have contributed to and approved the manuscript.

Additional information